A Method for Ontology and Knowledgebase Assisted Text Mining for Diabetes Discussion Forum

by

Ahmad Issa

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Engineering

University of Warwick, WMG

May 2015

Table of Contents

Acknowledgments

I am indebted to many people for the help and advice during my PhD journey, and I would like to thank them all for the support they gave me in their own, unique ways.

I would like to extend my deepest gratitude first and foremost to my academic supervisor, *Dr. Jay Bal*, who gave me endless support, patience and guidance throughout my research endeavour on both academic and personal levels, and shared with me his experience on academic life.

I am thankful for the financial support of the Higher Institute of Business Administration in Syria during my doctoral studies at University of Warwick, without which this research would not have been accomplished.

I would like to express my appreciation and love to my wife, *Dr. Issam Dayoub*, for her endless love and encouragement during my doctoral research, and for being always there for me. I thank her for believing in me and in my doctoral work. Without her constant support, this thesis would not have been possible.

I am especially grateful to my family without whom I would not be who I am today. To my parents, Mohsen Issa and Mariam Mostafa, whom I cannot express my love to in words, and my brothers for their unconditional love and encouragement.

Last but not least, I am also grateful to all my friends for their support and great companionship during my years at University of Warwick. *Dr. Majd Khador* for the great times we shared together and for being always there for me; *Dr. Ali Suker* for being a confidante and helping me to keep things in perspective; *Dr. Xiao Ma* for his help on this research and proofreading as well as for the useful discussions; *Mr. Radu Sora* for the generosity, encouragement and the helpful discussions along the way ; *Mr. Mohammad Seyyed Nabavieh* for the kindness and friendship; and many others of friends and staff who I will keep in my heart and memory.

Declaration

I, Ahmad Issa, declare that this thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy in Engineering. It has been composed by myself and has not been submitted in any previous application for any degree.

Abstract

Social media offers researchers vast amount of unstructured text as a source to discover hidden knowledge and insights. However, social media poses new challenges to text mining and knowledge discovery due to its short length, temporal nature and informal language.

In order to identify the main requirements for analysing unstructured text in social media, this research takes a case study of a large discussion forum in the diabetes domain. It then reviews and evaluates existing text mining methods for the requirements to analyse such a domain. Using domain background knowledge to bridge the semantic gap in traditional text mining methods was identified as a key requirement for analysing text in discussion forums. Existing ontology engineering methodologies encounter difficulties in deriving suitable domain knowledge with the appropriate breadth and depth in domain-specific concepts with a rich relationships structure. These limitations usually originate from a reliance on human domain experts.

This research developed a novel semantic text mining method. It can identify the concepts and topics being discussed, the strength of the relationships between them and then display the emergent knowledge from a discussion forum. The derived method has a modular design that consists of three main components: The Ontology building Process, Semantic Annotation and Topic Identification, and Visualisation Tools. The ontology building process generates domain ontology quickly with little need for domain experts. The topic identification component utilises a hybrid system of domain ontology and a general knowledge base for text enrichment and annotation, while the visualisation methods of dynamic tag clouds and co-occurrence network for pattern discovery enable a flexible visualisation of these results and can help uncover hidden knowledge.

Application of the derived text mining method within the case study helped identify trending topics in the forum and how they change over time. The derived method performed better in semantic annotation of the text compared to the other systems evaluated.

The new text mining method appears to be "generalisable" to other domains than diabetes. Future study needs to confirm this ability and to evaluate its applicability to other types of social media text sources.

Abbreviations

BOW	Bag of Words
CFM	Concepts Frequency Mean
CRM	Common Reference Model
CTV3	Clinical Terms V3
I.I.D	Independent and Identically Distributed
IE	Information Extraction
IR	Information Retrieval
ISI	Information Sciences Institute
KBTA	Knowledge-Based Temporal Abstraction
KD	Knowledge Discovery
LCHF	Low Carbohydrate High Fat
LOL	Laughing Out Loud
MERS-CoV	Middle East Respiratory Syndrome CoronaVirus
NCDs	Chronic non-communicable diseases
NE	Named Entity
NHS	National Health Service
NLG	Natural Language Group
NLM	National Library of Medicine
NLP	Natural Language Processing
OBIE	Ontology-Based Information Extraction
OGMD	Ontology for Glucose Metabolism Disorder
OGSF	Ontology of Genetic Susceptibility Factor
OOIE	Ontology-Oriented Information Extraction

PMI	Pointwise Mutual Information
POS	Part-Of-Speech
RCIF	Reduced Carbohydrates, Increased Fats
RDF	Resource Description Framework
SEA	Semantic (S) relatedness oriented ontology engineering via retrieving information from the search Engine (E) index with assistance from social network Analysis (A)
SML	Supervised Machine Learning
SMOB	Semantic Microblogging framework
SNA	Social Network Analysis
SNOMED CT	Systematized Nomenclature of Medicine Clinical Terms
SNOMED RT	SNOMED Reference Terminology
SOAP	Simple Object Access Protocol
TF-IDF	Term Frequency/Inverse Document Frequency
UMLS	Unified Medical Language System
URI	Uniform Resource Identifiers
VSM	Vector Space Model
WHO	World Health Organisation
WSD	Word Sense Disambiguation
WWW	World Wide Web
XML	eXtensible Mark-up Language

List of Figures

List of Tables

Chapter 1: Introduction

> "Reports that say that something hasn't happened are always interesting to me, because as we know, *there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns -- the ones we don't know we don't know*. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones." Donald Rumsfeld at a Department of Defence news briefing, February 12, 2002

Both *known unknowns* and *unknown unknowns* are of interest for scientific research. The researcher at the beginning does *not know* what outputs a research study will reveal, but often *knows* that the results are within a range of options – known unknowns. However, sometimes the results are totally unpredicted, when they reveal new unexpected knowledge – the *unknown unknowns*.

The focus of this research is about looking for the known unknowns, but also help uncover the unknown, unknowns hidden in social media web sites using knowledge discovery methods. The new *knowledge*, in this sense, is "not abstract"; it is based on common everyday human experiences (Ahmadabadi et al., 2006).

With the availability of huge amounts of raw information in social media sites, knowledge discovery methods can be used to help uncover previously unknown patterns in the stored data resources (Hardoon and Shmueli, 2013). Social Media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content (Kaplan and Haenlein, 2010).

Currently, social media, such as in blogs and social networks are the main source for unstructured data. They allow people to have a *voice* reaching out to the rest of the online world and provides a channel to share their information and personal experiences with others online. For example, social media empowers people suffering from specific illnesses (patients) as it facilitates interactions between patients to discuss conditions, share experiences and provide support (Sarasohn-Kahn, 2008). Patients often start off as information seekers when participating in a related online community, but often then a communication process starts among the patients, where participants share technical knowledge, emotional support, and personal experiences.

1.1 Research Motivation

Over the last twenty years, individuals' lifestyle and habits have changed significantly. These changes, such as unhealthy diet, physical inactivity and increased alcohol consumption, have resulted in a significant shift regarding the focus of healthcare systems shifting from communicable diseases to chronic non-communicable diseases (NCDs) such as cancer, diabetes, and cardiovascular diseases. The World Health Organisation (WHO) identified that more than thirty six million people are killed by non-communicable diseases every year (WHO, 2013b).

The social and economic costs of non-communicable diseases far exceed the direct treatment costs. NCDs affect individuals as well as the national health systems and the economy through a variety of mechanisms such as low productivity and high medical costs. There is an increasing burden on medical services, and a rise in the healthcare expenditure across the world in both developed and developing countries (Deloitte, 2014). Health expenditure in the UK for example, has grown from £54.6 billion in 1997 to £144.5 billion in 2012 (Figure 1-1). This rise in costs is driven by an aging population with a higher level of chronic diseases include cardiovascular, strokes, cancer and most importantly diabetes.

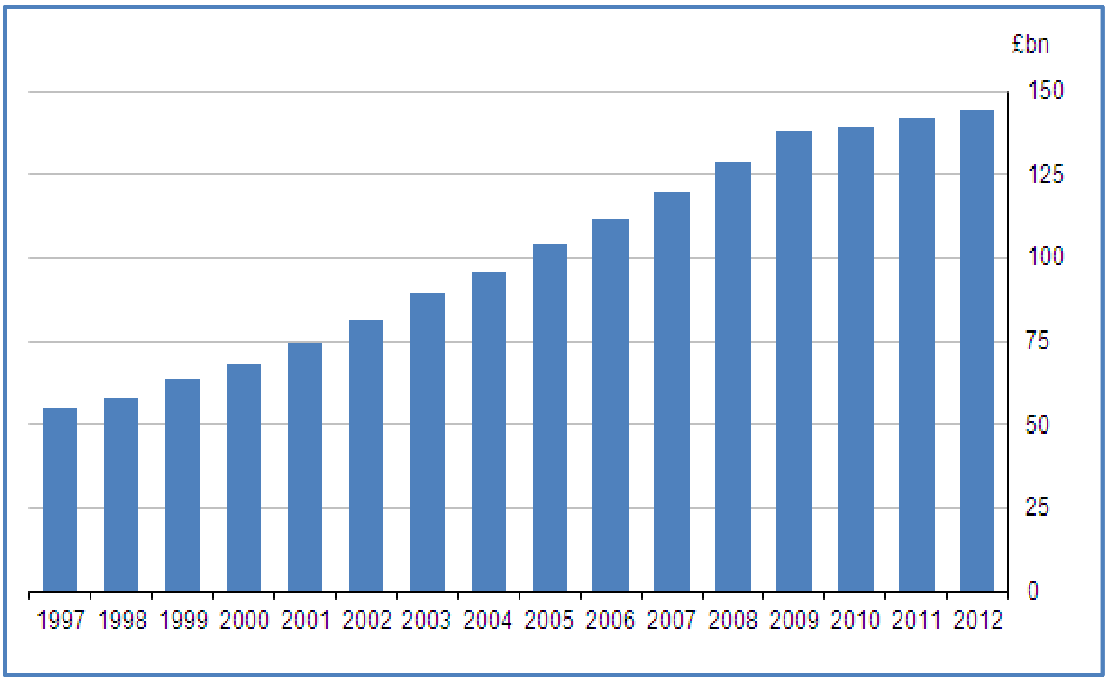


Figure 1-1: UK healthcare spending 1997-2012, source: (ONS, 2012)

Diabetes is considered as one of the great healthcare risks of the twenty first century, as the World Health Organisation (WHO) reported in 2013 that more than 347 million people had diabetes worldwide, which is a 78.86% increase in the number of people with diabetes globally in ten years period. The WHO also predicts that diabetes will be the fifth leading cause of death by 2030 worldwide (WHO, 2013a). The rise in diabetes can be seen as a worldwide challenge, it is not just an issue in the developed world (Figure 1-2). In western countries, it is estimated that between 5 and 11% of the healthcare budget is dedicated to diabetes, which (according to WHO) will increase in the coming years as the number of diagnosed people increases.

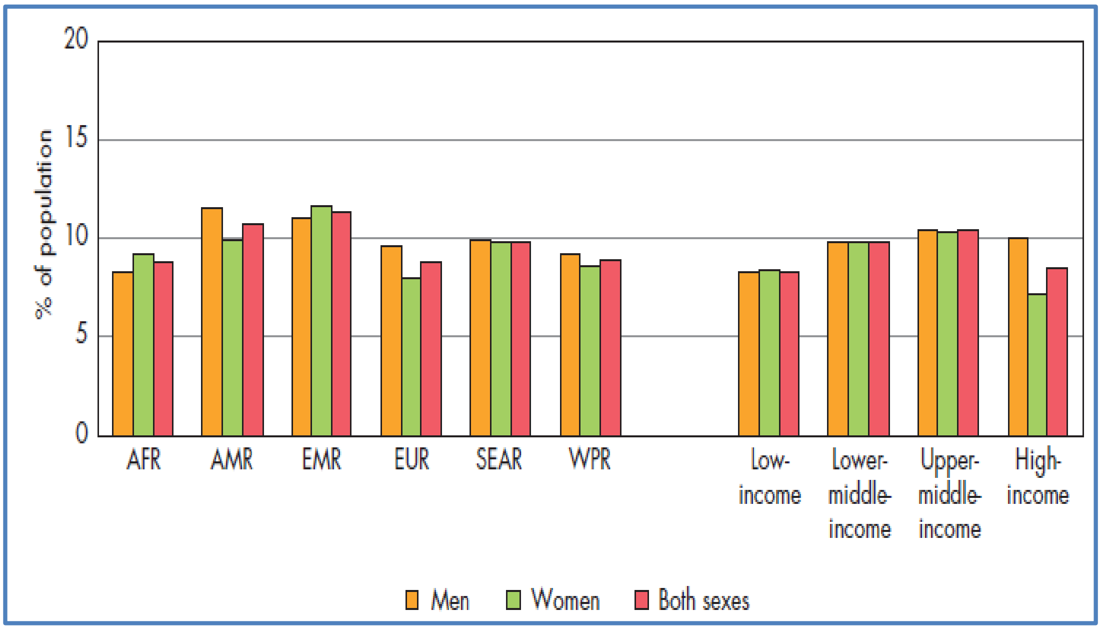


Figure 1-2: Age-standardized prevalence of diabetes in adults aged 25+ years, by WHO Region and World Bank income group in 2008, Source: (WHO, 2010)

In UK, there were more than 3.2 million people diagnosed with diabetes in 2013 (Table 1-1) and this number is predicted to increase to more than 5 million people by 2023 (Diabetes UK, 2014).

Table 1-1: Diabetes Prevalence in UK 2013, Source: (Diabetes UK, 2014)

Country	Prevalence	Number of Diabetics
England	6.0 per cent	2,703,044
Northern Ireland	5.3 per cent	79,072
Scotland	5.6 per cent	252,599
Wales	6.7 per cent	173,299
Total		**3,208,014**

The direct cost of diabetes in UK was around £9.8 billion in 2010-2011 (£1 billion for Type1 and £8.8 billion for Type2), which is around 10% of the total National Health

Service (NHS) budget. These statistics highlight the significant burden diabetes has on societies and healthcare services.

Several studies have shown that diabetes patients and their families are actively searching for information online about diabetes, for different purposes, such as treatment, diet and support (Zhang and Zhao, 2013; Greene et al., 2011). The successful management of this disease requires constant and dedicated efforts from both patients and healthcare professionals. With this typical chronic disease, it is essential for patients diagnosed with diabetes to become their own health managers and treatment experts. That requirement encourages patients to participate and engage in online forums and communities. In fact it seems that diabetes forums have become a significant source of information and support for diabetes patients (Zhang et al., 2013).

To assist communication between patients and medical professionals and to help uncover any hidden knowledge that the interactions and experiments going on within these forums may generate, it is important to try and analyse the discussions of sufferers on such online forums. This knowledge may emerge from consensus in the topics under discussion and arise from the many lifestyle experiments that are conducted and debated by users of the forums. These experiments may be related to diets, therapies and general lifestyle. For example, the use of music therapy and the best music to help cope with the stress of diabetes, as well as experimentation with different types of diets were important subjects of discussion on the main UK diabetes online forum.

Diabetes was chosen as the target domain for this research since it affects a significant percentage of the population. In order to explore the communication between patients, and to help define the challenges in analysing online discussion forums, a case study method was chosen. An empirical case study of the Diabetes Forums was used to explore, develop and test techniques for mining the text in these forums to help reveal information about what patients are really doing and experiencing as they try to cope with living with diabetes.

The data source for this case study was the Diabetes.co.uk forum[1], as it is one of the biggest diabetes online communities in UK (Table 1-2). It focuses on providing support and sharing experiences and suggestions among its members, and has more than 116,000 members who have posted more than half a million messages on the forum[2].

Table 1-2: Top Three Diabetes Forums in UK

Forum Name	Threads	Posts	Users
Diabetes.co.uk Forum	48,893	568,124	116,135
Diabetes Support (diabetessupport.co.uk)	46,380	506,262	12,645
DSF (diabetes-support.org.uk)	5,151	78,205	1,539

The discussion threads from this forum were downloaded, from its start in March 2007 till March 2013 (six years). The extracted text was used to test the text mining method developed by this research to help uncover the *known unknowns*, and reveal *unkown, unknowns*. The enhanced text mining method developed was specified by the requirements for text mining a health forum such as Diabetes.co.uk.

[1] http://www.diabetes.co.uk/forum/
[2] Forum Statistics was updated on 31/08/2014

1.2 Problem Description

The content of social media in general, and online discussion forums in particular, could be considered multi-faceted information. For example, a thread or a post in a diabetes forum might contain information about the diabetes type, symptoms, diet, or medication. Therefore, the collection of posts within the forum may have different relationships across these facets. The exploration and analysis of these topics and facets within online forums is a challenging task because of the distinct features of text posted in these forums, such as its temporal or unstructured characteristics.

In order to analyse the text in social media and extract useful patterns from it, text mining methods have been applied in different disciplines to uncover new knowledge from the unstructured text.

Text mining (or Text Analytics as it is sometimes referred to) is the analysis of text data so new knowledge can be discovered (Stavrianou et al., 2007) through automatic extraction of information from the text. An important part is connecting the extracted information to create new hypotheses that can be subsequently tested by traditional research methods, such as experimentation. In this sense, text analytics can be differentiated from other areas, such as information retrieval: In the case of information retrieval, the user is usually searching for the already known, something that has been written and indexed already. Hence, the challenge becomes finding the relevant materials according to the needs of the user, while text mining is different, since its purpose is to identify, classify and analyse unknown information from the text data.

However, traditional text mining methods, for example the ones that rely on the Vector Space Model (VSM) or the Bag of Words (BOW) model, have three main issues:

1. Mutual Independence Assumption: The use of a BOW model to represent the text results is a main limitation inherited from the assumptions of this model, where all the dimensions of the text are assumed to be independent from each other. This assumption considers all the terms and words in the text as mutually independent. However, the majority of terms and words in the discussion forums text are related - in one way or another - to each other. For example, the BOW model treats the words "Treatment", "Cure", and "Lantus[3]" as independent, although they are related. This is a key limitation of this model for text representation.
2. Lack of Meaning: Traditional text mining methods do not take into consideration the semantic features of terms and words in the text, such as hypernyms and hyponyms. For example, these methods treat "Basal Insulin" and "Lantus" as independent terms though "Lantus" is a brand name of "Basal Insulin" category. This may lead to having low scores of relevance between documents or posts even though they are highly related because they do not always use the same terms to describe similar ideas. This limitation fundamentally originates from the point that traditional text mining methods do not take into consideration the "meaning" of the analysed terms.
3. Dimensionality Curse: Traditional text mining based on a BOW model tends to utilise all the terms and words in the text (after stop words[4] removal). This might lead to hundreds of dimensions in text representation, and is known as the "Dimensionality curse". However, studies on text mining, such as by Yoo et al. (2007) Sobol-Shikler (2012) and (Shalaby et al., 2014) have shown that

[3] Lantus is a drug commonly prescribed for diabetes.
[4] Stop words are words that might be deemed general and meaningless, such as "*the*", "*are*" and "*is*".

a limited number of terms in the text can become the main elements for performing tasks such as clustering and summarisation. These terms have a distinguishable ability on analysing text and are usually the domain specific concepts in the text.

It has been demonstrated in multiple studies that text mining can benefit from the use of domain ontology and background knowledge to tackle the limitations in traditional text mining methods (Hotho et al., 2003; Bloehdorn and Hotho, 2006; Ranwez et al., 2013). A Domain ontology could be defined as the "formal representation of knowledge as a set of concepts within a domain using a predefined vocabulary to describe different types and relationships between concepts" (Kruse et al., 2015). Another way of defining it would be to consider the ontology as a set of groups or categories of interest and the relations between them for a specific domain. The grouping of concepts and their relationships could be seen as a construct to be understood by both humans and machines.

For example, Bloehdorn and Hotho (2006) proposed an approach to incorporate background knowledge into a text mining system for text classification purposes. They demonstrated through three different experiments that their approach improves text classification results compared to the "Bag of Words" model. However, the results have depended on the ontology chosen for providing the background knowledge.

In text mining systems for specific applications (especially the ones with some noticeable domain explicitness), information extracted from external domain knowledge sources could be utilised to improve key elements in the text mining process significantly, such as text pre-processing, text representation, or the knowledge discovery tasks (Feldman and Sanger, 2007). Therefore, the creation of suitable domain ontology is important for finding the known and the unknown, unknowns in an effective text mining approach.

However, building a suitable ontology tailored for the needs and requirements of the text mining process is still a major problem and an important area of research. Locating a suitable ontology (especially in multidisciplinary areas) is often a difficult task. It either does not exist or if it does, the work required to adapt it to the text analysis requirements is significant. A cause of the need to make domain representation modifications may be rapid changes in the domain knowledgebase and structure. For example, an ontology for, say, the IT domain which is five years old may have significant gaps when using it as a support knowledge base to a text mining application in the IT field. This issue in the text mining process would lead to poor extraction and analysis of information and poor performance from the semantic text mining system that uses these types of ontology.

In addition, existing ontologies are usually built according to the requirements of their applications, making it harder to re-use and customise them to be suitable to other applications (Barrios and Vilches-Blázquez, 2010). These issues lead to a focus on automatically generating domain ontology to support text mining for this research.

As the text in non-expert domain specific social media content (such as medical forums) has a high level of multi-disciplinary content and multiple facets, the mining of the discussion forums requires exploring and examining methods to tackle its complexity through the use of background knowledge. This becomes a key area of investigation for this research.

1.3 Research Question

The research question for this research is:

“How can we identify and display the concepts and topics being discussed, the strength of the relationships between them and emergent knowledge in online discussion forums?”

Based on this question, three main sub-questions were identified:

1. How can we identify the concepts in a particular domain and the relationships between them? This involves not only concepts at an expert level, but also at a novice level and linking these two.
2. How can we understand the emergent knowledge behind text conversations?
3. How can we present the living nature of an online discussion forums with themes dying, being reborn or new ones emerging?

The chosen methodology to answer these questions is one of creating a novel text mining method for analysing text in social media. In order to generate such a method, three objectives have been identified:

- To examine and discuss current text mining problems when analysing social media;
- To devise methods to tackle these problems and identify the emergent knowledge and important insights behind the text conversations, and evaluate these methods;
- To discuss the generalisability of the proposed approach.

Included in this thesis is the methodology for deriving the text mining approach and evaluating it using a diabetes discussion forums case study, where methods are described in the context while they are required due to the nature of the developed approach and in order to make a coherent and comprehensive presentation of the derived approach.

1.4 Outline of the Work

The rest of the thesis is organised as follows:

Chapter two reviews the related work on methods for text mining and analysis in social media. Particularly, it discusses the traditional text mining methods and their limitations when applied to social media, and the role of background knowledge in text mining to address those limitations.

Chapter three describes an automatic ontology building process that uses a knowledge source to obtain domain specific terms and relationships, with little need for domain experts. This ontology must be suitable for enriching and annotating the text for further analysis.

Chapter four proposes a text mining and visualisation method based on a hybrid system of domain ontology and general knowledgebase for semantic annotation and topic identification. It also extends tag cloud visualisation and describes a new dynamic tag cloud method. The text mining method utilises a co-occurrence network to identify interesting relationships between concepts and topics, and to uncover hidden relationships.

Chapter five describes the results of applying the text mining method derived to the diabetes case study and the evaluation of its effectiveness. The diabetes ontology for this research is produced/derived and evaluated from content and practical perspectives. Then, the text in the diabetes discussion forum is analysed and the results are visualised using a novel derived method. An evaluation of the derived text mining method is conducted to examine its generalisability to other domains.

Chapter six provides a summary of the work and discusses the contribution to knowledge. The text mining method derived provides a novel ontology building process, an automated semantic annotation method and visualisation tools for text analysis. Finally, this chapter discusses the limitations of the research and the future opportunities arising from this research. The structure of the thesis is illustrated in Figure 1-3.

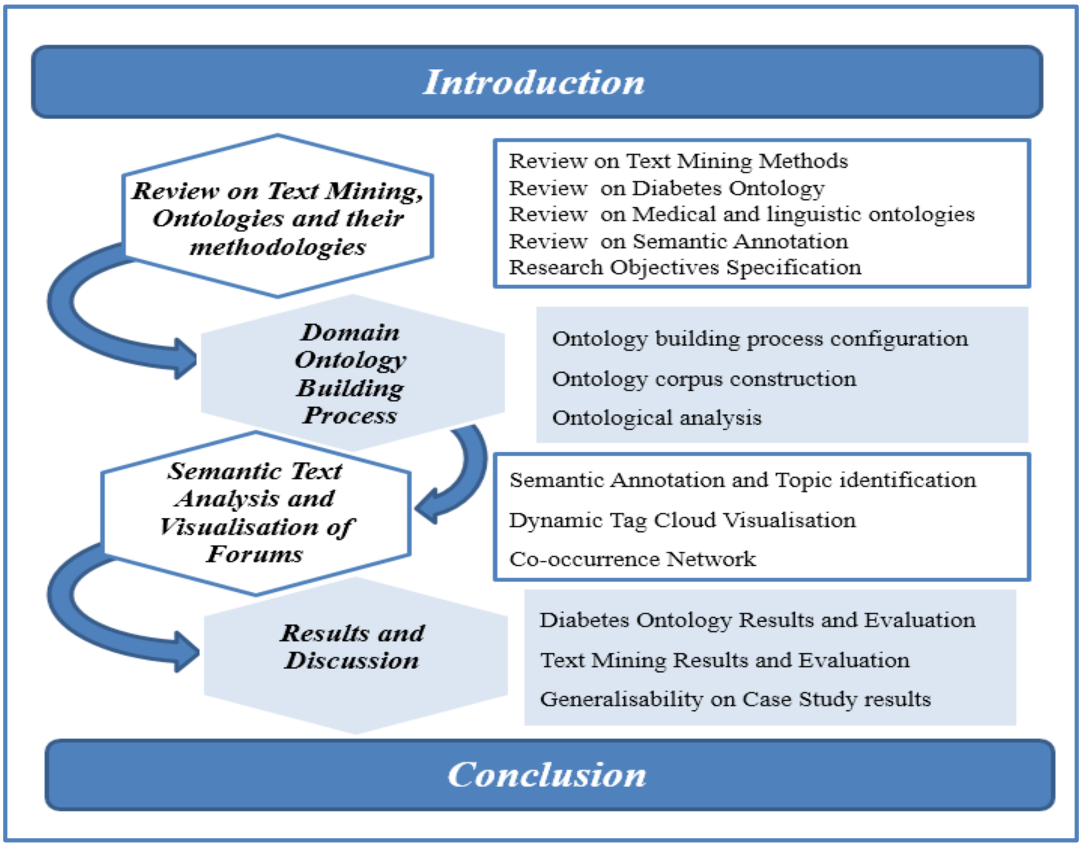


Figure 1-3: Thesis Structure

Chapter 2: Text Mining of Online Discussion Forums

The goal of this research was to investigate methods to analyse text extracted from online discussion forums in order to reveal hidden knowledge and identify concepts and their relationships in the discussed topics.

This chapter discusses the traditional text mining techniques and reveals how such techniques may not fully satisfy the text mining requirement in social media due to the distinctive characteristics of the text data within it. Then, the current approaches suggested in the research literature to address these characteristics are discussed.

2.1 General Text Mining Architecture

From an abstract perspective, a text mining process takes in raw text as an input and produces different kinds of output, such as patterns, trends or relationships maps. Figure 2-1 demonstrates this simple paradigm.

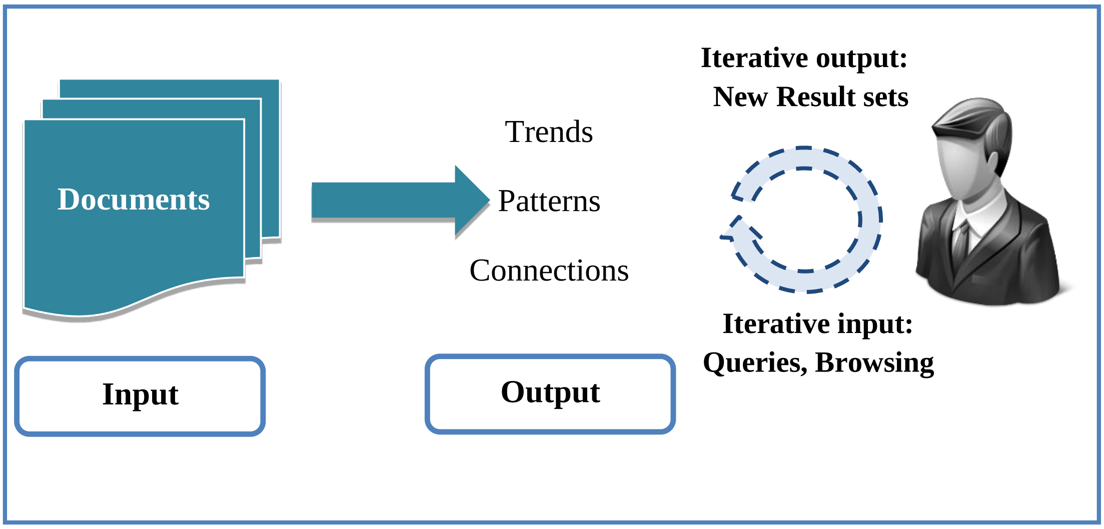


Figure 2-1: Basic input–output model for text mining

On a functional level, traditional text mining architecture follows a general model that includes three successive stages (Feldman and Sanger, 2007): Text Pre-processing, Text Representation and Knowledge Discovery (K.D.). Figure 2-2 illustrates these steps and they are discussed in more detail in 2.1.1, 2.1.2, and 2.1.3.

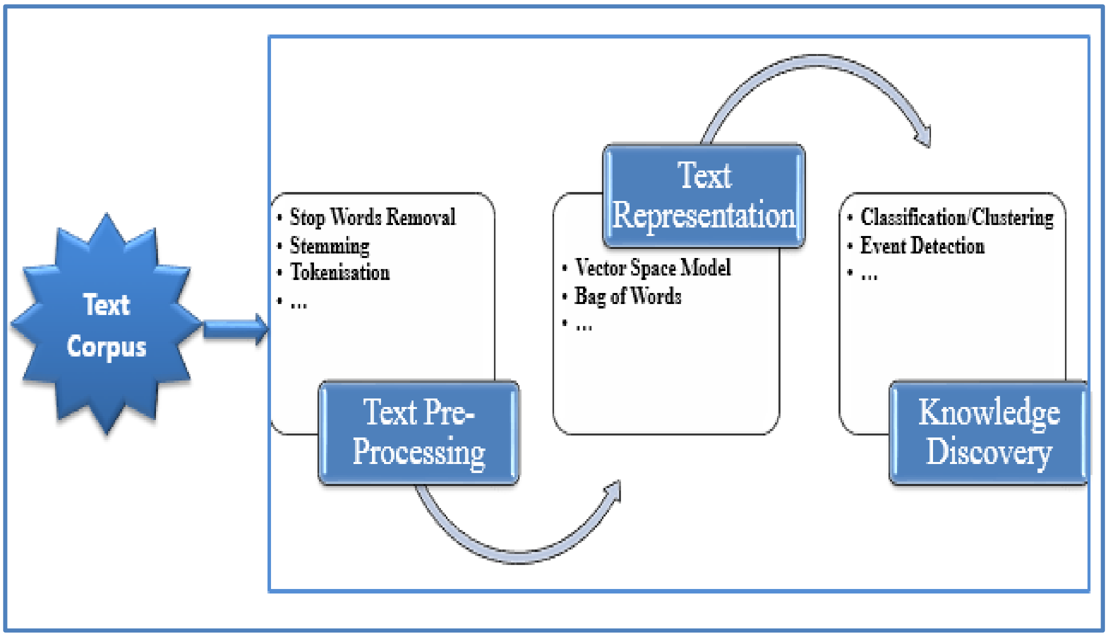


Figure 2-2: Traditional text analytics architecture

2.1.1 Text Pre-Processing

The goal of text pre-processing is to create consistency in the document (text corpus) in order to facilitate the next phase of the process - text representation. The pre-processing phase is employed in the majority of text analytics systems. Traditionally, this phase includes three main steps demonstrated in Figure 2-3:

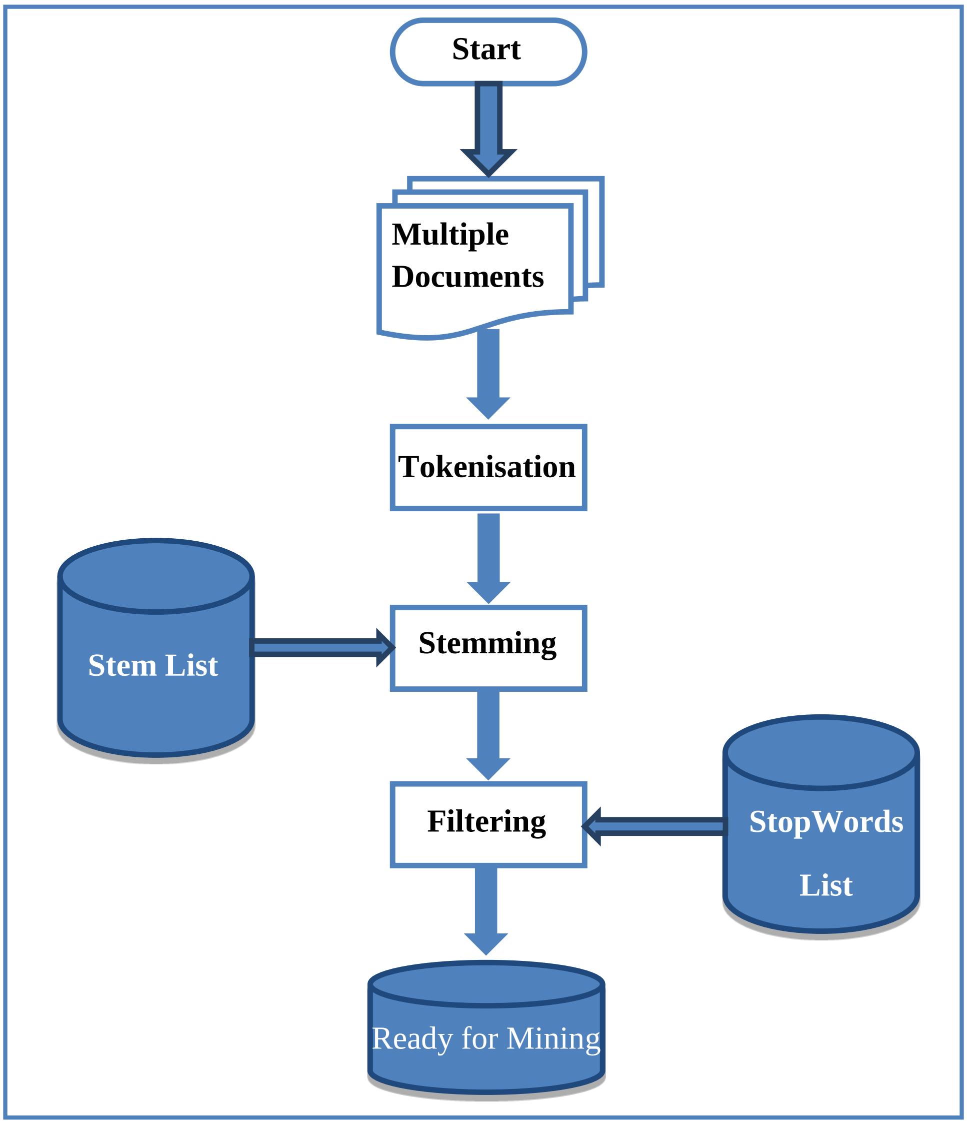


Figure 2-3: Text mining pre-processing phase

1. **Tokenisation**: This task splits the text into individual sentences and then splits the sentences into individual words (known as *tokens*). Figure 2-4 shows an example of text tokenisation. These tokens are then used in the later stages of text processing. The group of different tokens acquired from all the documents of a text corpus is known as the "Document collection dictionary".

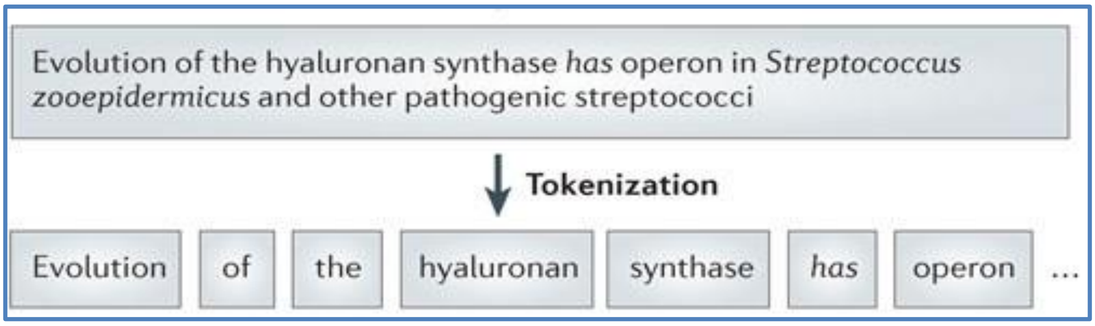


Figure 2-4: An example of text tokenisation

2. **Filtering**: This task removes specific tokens from the text using a pre-defined list. This filter list usually contains words that might be deemed general and meaningless, such as "*the*", "*are*" and "*is*".
3. **Stemming** (Porter, 1980): This process identifies the original "*root*" or "*stem*" form of the derived words. For example, "Walk", "Walking", and "Walked" have the same root: "Walk". This allows the text analysis process to regard words with variant forms as the same feature, which reduces the *dimensionality* of the document's description within the corpus. This approach is used in many information retrieval systems in order to enhance the matching between user queries and retrieved documents.

The text pre-processing phase varies based on the specific text analytics application. For example, in opinion mining or Natural Language Processing (NLP) applications, the textual analysis depends on the syntax of the text. This requires that the pre-processing method maintains/preserves the original structure of the sentence.

Often, the next text mining phases can be applied without any further text pre-processing. However, some applications require additional pre-processing in order to preserve the richness of information about words and concepts within the text. For example, part-of-speech (POS) tagging identifies and marks words based on their part of speech, i.e. noun, verb or adjective. It is often used as a pre-processing step in event detection applications to identify the verbs and nouns, which in turn are used to match a specific pattern of events in the later stages of the text mining (Hogenboom et al., 2010).

2.1.2 Text Representation

Current technologies cannot “read” or “understand” text as our brains do because of the unstructured nature of free text. Thus, text analytics creates different levels of structure in the textual data in order for mining algorithms such as machine learning to process the data by reducing the complexity of free text. For example, in the Information Retrieval (IR) domain, this problem has been tackled using vectors to represent documents. User queries and documents in IR have been modelled in vector space for decades (Srivastava and Sahami, 2010). This vector-based method tries to represent documents by converting them into numeric vectors, which in turn can be dealt with using linear algebraic operations, such as vector distance measures (Berry and Castellanos, 2008). This method is known as the “Vector Space Model” (VSM) (Salton, 1989), or “Bag of Words” (BOW). In VSM model, any word is represented using a separate variable that carries a numeric weight of importance (Singhal, 2001). Because any document includes a limited number of terms/words (out of all the possible words and terms in the domain), the majority of the resulting vectors are “sparse” numeric vectors.

Using this model for simple representation of the textual data, the linguistic structure of the text is lost or ignored. Figure 2-5 shows an example of the representation of unstructured text using vectors.

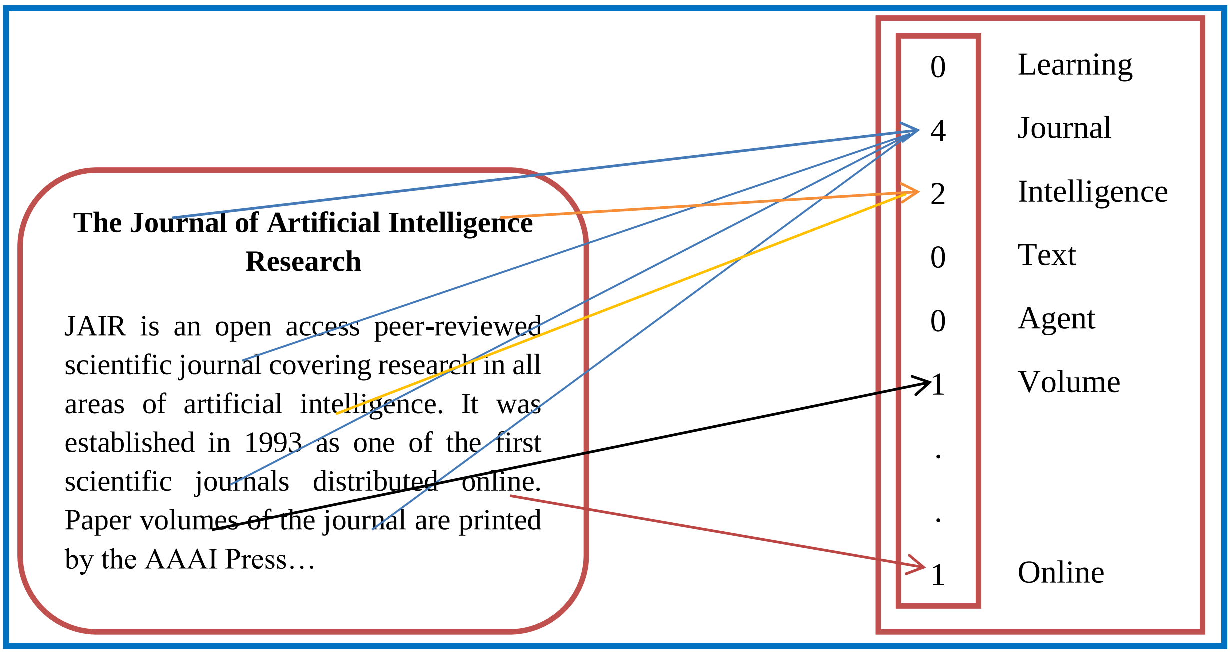


Figure 2-5: Example of BOW document representation

The most common way of calculating the weight of words in vector-based models is Term Frequency/Inverse Document Frequency (TF-IDF) (Manning et al., 2008):

$$\boxed{tfidf\,(\omega) = tf\,(\omega) * \log \frac{N}{df\,(\omega)}}$$

Equation 2-1

Where:

- $tf\,(\omega)$: is the word frequency (the number of occurrences in the given document);
- $df\,(\omega)$: is the document frequency (the number of documents that contains ω);
- N: is the number of all documents in the text corpus;
- $tfidf\,(\omega)$: is the relative importance of the word in the given document.

Figure 2-6 shows that by applying this model to three posts from twitter with TF-IDF weighting, the text corpus is represented as a matrix. Each row in this matrix represents a distinct word (six words in this example) and each column in the matrix represents a post, i.e. three documents (posts) in this example.

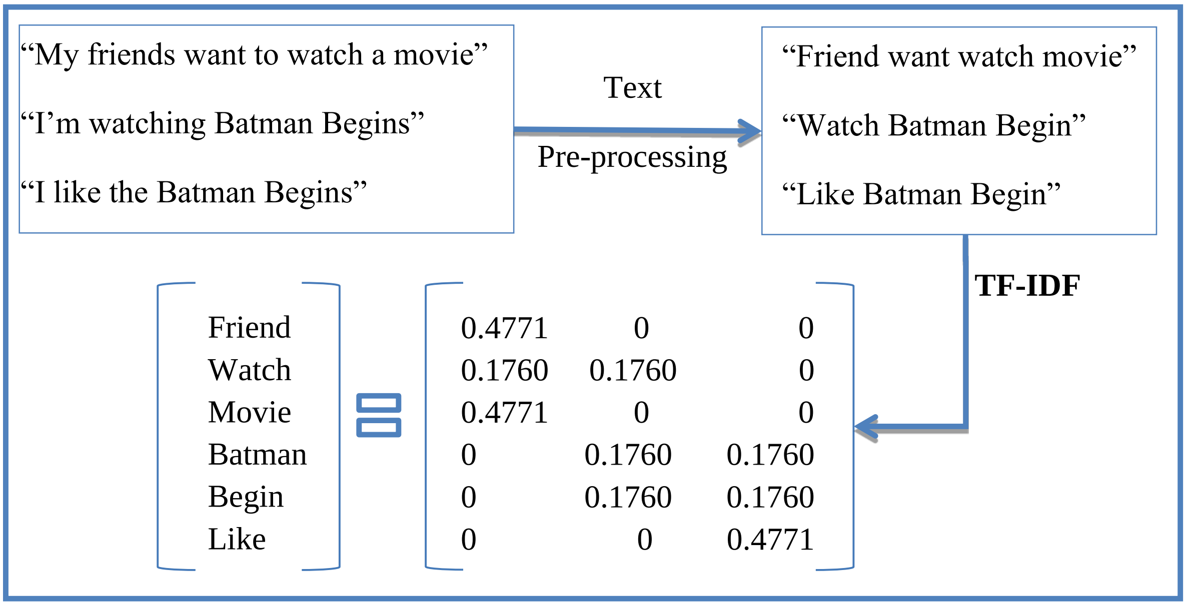


Figure 2-6: TF-IDF calculations for three social media posts

The output of this step can be then utilised in the knowledge discovery phase of the text mining system.

2.1.3 Knowledge Discovery

When the text corpus has been transformed into vectors using the VSM/BOW method, data mining/machine learning methods and algorithms such as clustering or classification can be applied. An example of a knowledge discovery method is the "Similarity" measure widely used in many machine learning tasks. The most successful way of calculating similarity between two document vectors D1 and D2 is "**Cosine Similarity**" (Nyein, 2011):

$$\mathrm{Sim}(D_1, D_2) = \cos(\theta) = \frac{D_1 . D_2}{\|D_1\| \, \|D_2\|} = \frac{\sum_{i=1}^{n} D_{1i} \text{ x } \sum_{i=1}^{n} D_{2i}}{\sqrt{\sum_{i=1}^{n} (D_{1i})^2} \text{ x } \sqrt{\sum_{i=1}^{n} (D_{2i})^2}}$$

Equation 2-2

In the case of text matching applications, the vectors D1 and D2 usually represents the term frequency vectors of the associated documents. The cosine similarity measure can be regarded as a way of normalising the length of documents while comparing them. In this case, the similarity value ranges between -1 and +1, where -1 means that the documents are exact opposite, while +1 means they are identical.

In other applications, such as information retrieval, D1 and D2 would usually represent the TF-IDF weights from text representation phase, which are not negative. This means that the cosine similarity between the two documents will range between zero and one.

Utilising traditional text mining methods, such as the Bag of Words model, researchers can obtain useful information from the text corpora such as the similarity between two documents. However, this model faces several challenges when used directly to mine text extracted from social media/forums due to the distinctive characteristics of this data.

2.2 The Distinctive aspects of Text in Social Media/Forums

The textual data in social media has distinctive characteristics when compared with traditional documents. These characteristics are (Bontcheva and Rout, 2014; Aggarwal and Zhai, 2012b):

1. Temporal Nature of Text;
2. Text Length;
3. Unstructured Nature of Text;
4. Rich Information Content.

There follows a discussion of each one of these characteristics.

2.2.1 Temporal Nature of Text

The temporal nature of text in social media platforms is a very important, common and distinct feature. In particular, users of microblogging and social networks sites may post news or information many times in a single day, while bloggers often update their blogs several times a week. As well as sharing and communicating with other friends and followers, users of social media comment on the most recent events that they consider important to them, such as sports, political events, news, movies, or new products. These posts and comments have a time-stamp associated with them. Hence, the text mining of these posts and comments should take into consideration the temporal nature of this data. According to Bontcheva and Rout (2014), addressing this dimension of textual data in social media is still a relatively "under-researched problem".

More importantly, with the fast developing nature of the styles of social media content and communication, the text itself is changing too. Traditional machine learning approaches often suppose that text data are *independent and identically distributed* (I.I.D) (Hu et al., 2013). However, data can often be split into subgroups, which are related in one way or another. This violates the original assumption that the data points are identically distributed. These subgroups are known as domains, and the data in these domains is not considered independent and identically distributed as traditional machine learning approaches assume. Therefore, the text mining models that are based on these methods are not suitable for analysing social media text.

The temporal nature of text in social media presents a significant challenge to text mining activities. Also often, this text is short in length and thus does not contain redundancy, which can be useful in text analysis.

2.2.2 Post Length

Most social media platforms limit the size/length of the users' messages and posts. For example, Twitter provides users the ability to share news swiftly, but limits the length of each tweet to hundred and forty, UTF-8 encoded characters. Another example is the photos sharing platform Picasa (picasaweb.google.com) which limits the captions to 1024 characters and comments to 512 characters.

The short-length nature of the posts and messages on social media make them more efficient (in term of time and effort) while engaging in social media discussions (Aggarwal and Zhai, 2012b). However, this feature presents new challenges for traditional text mining tools and techniques. Unlike traditional text corpus, short posts on social media sites contain fewer sentences, which means they are not suitable to be analysed using traditional methods such as BOW (discussed in 2.1.2), as then these techniques do not achieve high accuracy because of data sparseness. For example, short messages may not provide enough contextual information to perform knowledge discovery similarity calculation (Phan et al., 2008).

2.2.3 Content Quality and Unstructured Sentences

The quality variance of the textual content is an important difference between social media text and traditional sources. This variance stems from two origins:

I. The users' attitude towards the topic affects the quality when commenting on a post in a forum or posting a tweet. Some social media users are experts in the topic and consequently they post very carefully with high quality content, while other users might be learners to the subject/topic and thus their posts do not have the same quality level. The main problem that the text in social media presents is the high variance of the quality distribution from low-quality posts to very high-quality ones. This complicates the tasks of ranking and filtering in textual analytics systems compared with other domains (Agichtein et al., 2008). In addition, some posts are considered “noisy” as they are not related to the topic discussed. For example, one online forum user’s post says “I like Apple products because they have high quality” can be considered noisy data if the topic is “faulty brakes in Toyota vehicles”. Thus, it is hard to classify and filter such a post into the correct class/cluster without taking into consideration the context information of the discussion.

II. Social media users, when writing a post or a comment might use new acronyms or abbreviations, which cannot be captured using traditional methods since they do not appear in the usual text documents. For instance, the word “LOL” is a very popular internet slang acronym for “laughing out loud” and is used very intuitively in social media, but it was not used in conventional text documents previously. Such acronyms give social media users convenience while communicating and sharing with friends. It is however very difficult for automatic systems to capture accurately the meaning of the messages that contains these abbreviations.

2.2.4 Rich Information Content

In general, social media demonstrates a wide range of information sources. For instance, the hashtag symbol “#” used in Twitter (and in Facebook from December 2013), is utilised by users to mark specific keywords or topics (information tagging). With the explosion of digital photography, images are usually tagged with multiple classes or labels that are symbolised by different areas in the photo (Zha et al., 2008). One may usually find a blog or a microblog post with associated links included in the post, especially in limited-length platforms such as Twitter. Another example of this is Wikipedia, which provides an effective method for users to be redirected to the "*Disambiguation Pages*" or higher concept pages. This associated rich information presents great opportunities for information extraction. The textual analysis of social media corpus can utilise data from a variety of aspects such as text, hashtags, or time stamps.

For example, Wang et al. (2010) used users-generated tags to determine the "overlapping communities" in a blogosphere. They proposed a clustering model that takes advantage of network structure and tags in the social media in order to discover connected communities. In this model, tags and users were associated to each other, which was useful in detecting the evolution of the communities through observing the changing interests of users. Another example in Social Network Analysis (SNA), where researchers have utilised a variety of information such as tags and links to detect influential users in the blogosphere (Agarwal et al., 2008; Xie et al., 2013) and in other social networks sites, such as Facebook and Twitter (Bakshy et al., 2011; Grabowicz et al., 2012), to understand the behaviour of social media users at a group level (Kwak et al., 2010; Goh et al., 2013). In the knowledge management area, Lu et al. (2010) and Wyner et al. (2012) used the profiles of product review authors and other social context information from their social network in order to improve the automatic assessment of the review quality. This was superior to previous work, such as Ghose and Ipeirotis (2011), Liu et al. (2007), and Kim et al. (2006) that attempted to solve this problem by considering each review as an individual document and extracting features from the text for quality assessment of the review.

These four characteristics of text in social media have presented a big challenge to traditional text mining methods when trying to analyse social media text. In section 2.2.5, methods to bridge this semantic gap are discussed.

2.2.5 Bridging the Semantic Gap

Social media content is characterised as being short and unstructured. When traditional text analysis frameworks process this type of data, their approach is fundamentally limited, since they only use explicit information in the text (Gabrilovich and Markovitch, 2005; Aggarwal and Zhai, 2012a).

Two categories of approaches have been suggested in order to address the BOW sparseness issue (section 2.1.2). The first is a direct representation of the text known as a “Surface representation” (Kumaran and Allan, 2004) that analyse the sentences in the text from multiple facets and maintain the context of the text. This approach is time consuming for detailed structural analysis of the text (Aggarwal and Zhai, 2012b). Another limitation is the failure to exploit external knowledge, which has proved beneficial in addressing the “semantic gap” issue. For example, the tweet’ hashtag “Fukushima Earthquake” does not give any indication associated with it being a “nuclear disaster”, though it is easy to recognise that they are related by exploring other news sources. BOW-based methods (discussed in 2.1.2) find it difficult to build any semantically related connection between the two. Several approaches have been proposed to enhance basic text context through the use of external linguistic resources. Such approaches were found to be useful in reducing the “semantic gap” in a variety of text and data mining tasks (Dou et al., 2015).

In order to tackle the “semantic gap” problem, researchers have used domain knowledge and ontologies to try to bridge the gap faced while representing text in social media. It has been demonstrated in multiple studies that text mining can benefit from the use of domain ontologies and background knowledge (Hotho et al., 2003; Bloehdorn and Hotho, 2006; Ranwez et al., 2013). The use of ontologies and metadata produces high quality results, yet the cost of manually deriving domain ontologies and manually tagging the corpus is many times higher than using traditional methods. In the next section, the use of background knowledge and ontology in text mining is discussed.

2.3 Using Domain Knowledge for Text Mining

Concepts and terminology resulting from the "Text Representation" phase in the general text mining architecture (discussed in section 2.1.2) are not only included in the documents' attributes, but also belong to a specific domain. A domain can be defined as a particular, focused area of interest that can be represented by "lexicons, ontologies and taxonomies of information" (Feldman and Sanger, 2007). For example, domains can represent as broad areas of interests as in "*economics*", or narrower ones such as "*micro-/macro- economics*" or "*Mergers and Acquisitions (M&A)*".

In the classical literature on data mining, domain and background knowledge are used as an instrument to constrain knowledge discovery's processes. From the data mining literature three main kinds/types of the domain knowledge (provided by external sources) used by data mining applications can be generalised (Anand et al., 1995; Feldman and Sanger, 2007):

1. Relationships Rules,
2. Domain Knowledge Categories, and
3. Knowledge Discovery Constraints.

Since then in the last decade several studies in the field of text mining have proposed other implementations of domain knowledge that can be valuable for text mining applications (Silva and Ribeiro, 2007; Hotho et al., 2003). Information extracted from external knowledge sources could be utilised to improve key elements in the text mining architecture significantly, such as text pre-processing, text representation, or knowledge discovery tasks. Theoretically, even text mining systems that do not include clear domain orientation in their design may benefit from the addition of information extracted from external knowledge sources, which are related to broader domains, such as English lexicons, reference works and encyclopaedias.

Domain knowledge may be used in the processes of the “text pre-processing” phase of any text mining system in order to improve the extraction of concepts/terms, as well as the validation of the results. Moreover, the use of domain knowledge can be essential in the improvement and expansion of deep, reliable and controlled “Concept Hierarchies” (Feldman and Sanger, 2007). For example, one of the uses of the domain knowledge is the formation of significant constraints for the “Knowledge discovery” tasks, where these constraints are used to evaluate the patterns identified by the knowledge discovery process. In this case domain knowledge can be used to build specific constraints and enhance the results of text mining applications such as text clustering (Huang, 2008; Charola and Machchhar, 2013). Using the domain knowledge, several text mining techniques have been suggested in order to bridge the “semantic gap” and address the data sparseness problem in BOW-based systems (discussed in 2.2.5) with short text data. For example, Banerjee et al. (2007) suggested the use of Wikipedia as an external source for including additional information/features in order to improve the representation of short text data.

The integration of domain knowledge in text mining systems plays an important role in text analytics. Therefore, analysing diabetes discussion forums would benefit from utilising domain ontology and lexicons in order to address the semantic gap problem. Therefore, a discussion of the approaches to generate a domain ontology for the diabetes case study and its application in text mining is needed.

2.3.1 Domain-specific Ontologies and Lexicon

Several text mining frameworks have incorporated background knowledge bases in the shape of domain ontology or a domain lexicon. In order to classify concepts accurately, experts have a tendency to utilise specialist language in a specific domain. This tendency to be specific and accurate makes it difficult to construct background knowledge with the required level of “fuzziness” or redundancy to make it accessible to non-experts. Often the domain specific ontology used by text mining applications lacks the required fuzziness to be applied to real-world systems (Abulaish and Dey, 2007).

Several researchers have observed that online users’ tags, might offer the required broad connections to the relevant non-specialist concepts. The tags are the result of the online users’ personal tagging of information on the web and usually known as a “folksonomy” (Vander Wal, 2007; Carmel et al., 2012). This tagging is generally achieved in a shared and open social manner. However, because of the “free” nature of tagging, folksonomies tend to be messy and chaotic since any concept (i.e. tag) or relation from any viewpoint may enter into the folksonomy (Specia and Motta, 2007). Moreover, the large number of messy and chaotic terms and their relationships creates an obstacle to organising the domain-related concepts into subgroups (Xu et al., 2006; McKenzie, 2013). Therefore, folksonomies may not be appropriate to be the base for a domain-specific ontology or terminology for text analysis.

As the case study for this research is in the diabetes domain, previously developed diabetes ontologies may be able to contribute to diabetes forum analysis.

2.3.1.1 Diabetes Ontology

There have been several efforts from researchers and healthcare professionals towards developing an ontology for diabetes in the last twenty years. Each one of these ontologies was built for a specific application; hence, they prioritize different issues. In this section, diabetes ontologies are discussed for their suitability to analyse online discussion forums.

2.3.1.1.1 Knowledge-based temporal abstraction (KBTA)

Shahar et al. (1994) have proposed a general method for creating concepts from clinical data, called *knowledge-based temporal abstraction (KBTA)*, and used it for monitoring patients. This method has focused on knowledge acquisition and reuse. The case study used by KBTA was Diabetes mellitus type 1 (known also as *juvenile diabetes*) because one of the researchers who developed the ontology (Frederic Kraemer) was the domain expert (diabetologist). With his help, they created three ontologies: Parameter-properties, events and context ontologies. Figure 2-7 shows the context ontology. The proposed system was applied to analyse Type1 diabetes patients' records in order to monitor their status using different parameters such as the glucose level.

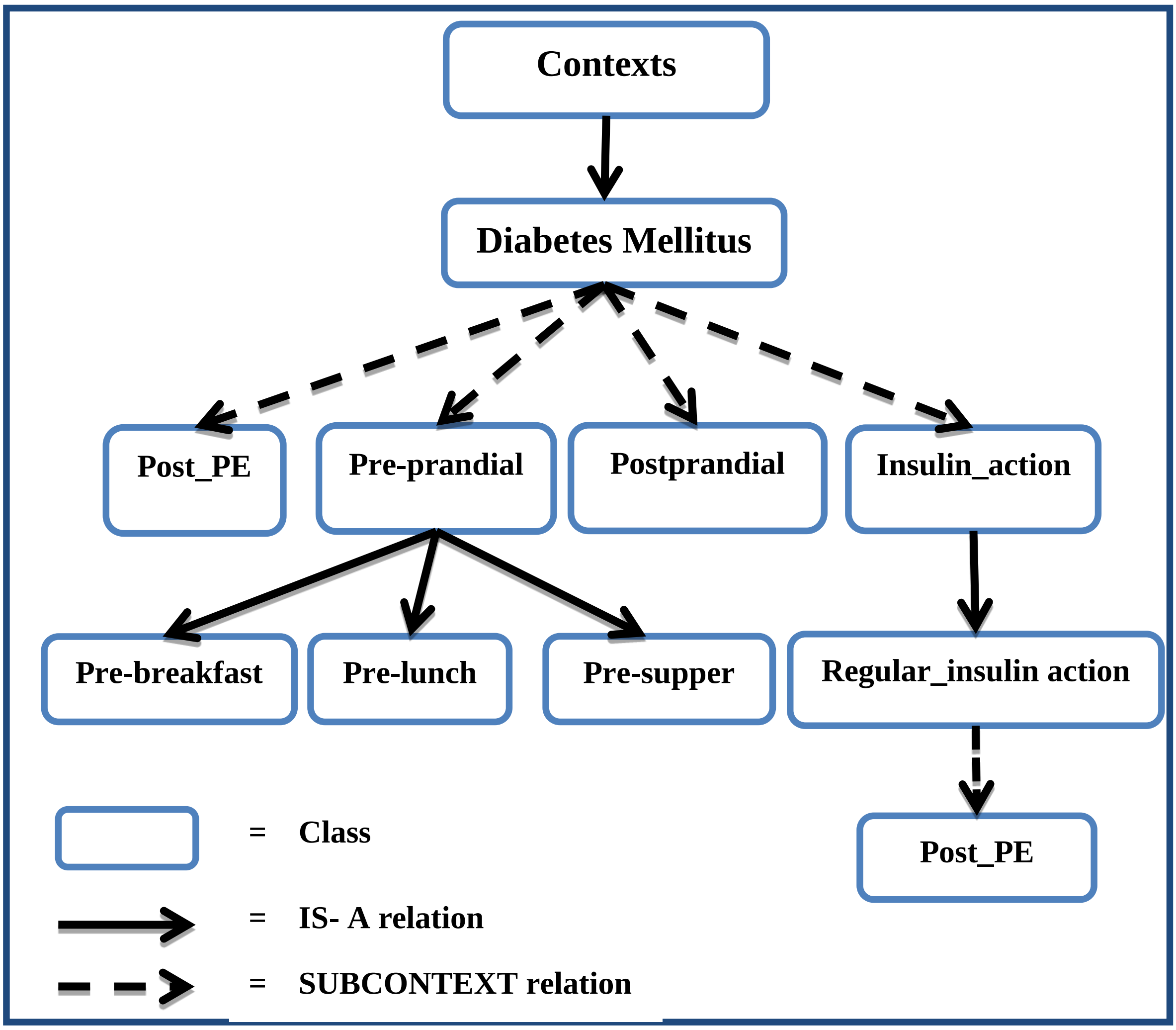


Figure 2-7: KBTA diabetes mellitus type1 ontology, source: (Shahar et al., 1994)

This ontology does not cover Type2 diabetes mellitus, which is the most common diabetes type and the most discussed in literature and between patients. In addition, the ontology provided insufficient quantity of terminology for the diabetes domain and has reflected the domain expert, not the user point of view in its development.

Ganendran et al. (2002) developed a multi-agent system for healthcare management based on a medical ontology and applied it to a diabetes case study. The system's goal was to improve the collaboration between medical professionals (doctors, clinicians and specialists). Figure 2-8 illustrates a simple model of the diabetes ontology used. Each class in the ontology defines a concept in the diabetes domain. Class *Diabetes* represents the main concept in the domain that contains sub-classes such as *Symptoms* and *Treatment*. Each class has a number of attributes associated with it. These concepts and the relationships model the diabetes ontology. The model was used to analyse and answer queries.

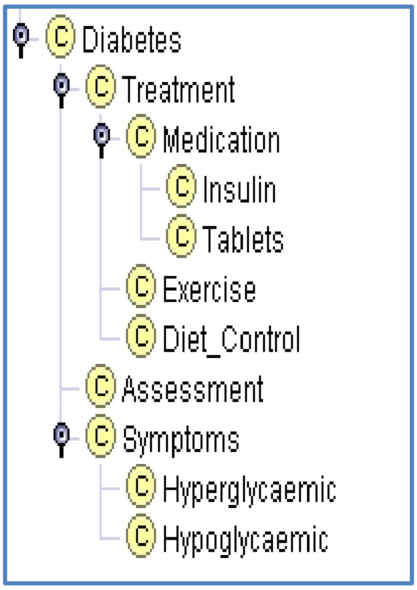


Figure 2-8: Simple diabetes ontology for multi-agent system, Source: (Ganendran et al., 2002)

Diabetes ontologies have also been used for creating knowledge management framework for knowledge modelling and acquisition.

2.3.1.1.2 Diabetes Healthcare Knowledge Management

Buranarach et al. (2009) introduced a semantic framework for chronic disease healthcare knowledge management. They stressed the importance of using an ontology for knowledge management systems in the medical domain, especially decision support system and clinical information systems. The case study in their research was to apply the framework to the diabetes healthcare domain. The derived diabetes ontology for their research was used as a tool for knowledge acquisition and modelling. The ontology was derived by domain experts based on the formal clinical guidelines in Thailand. Their diabetes ontology models two types of knowledge (Figure 2-9):

1- Structural Knowledge (Figure 2-9): This type allows the system to make use of patients’ records and, therefore, this knowledge offers the *schema* of patients’ data. This contains the historical data of assessment and therapy, which is vital for both decision support and clinical information system.

2- Procedural Knowledge: This process-oriented type represents the recommendations offered by the clinical guidelines and is essential for the decision support system guiding the medical processes of assessment and treatment. The combination of this knowledge with the structural knowledge helps medical caregivers make informed decisions regarding their patients.

```
<owl:Class rdf:ID=”DM_patient”>
  <rdf:subClassOf rdf:resource=”#user”/>
</owl:Class>
<owl:DatatypeProperty rdf:ID=”hasName”>
  <rdfs:domain rdf:resource=”#user”/>
  <rdfs:range rdf:resource=”&xsd;string”/>
</owl: DatatypeProperty >
<owl:DatatypeProperty rdf:ID=”hasAge”>
  <rdfs:domain rdf:resource=”#user”/>
  <rdfs:range rdf:resource=”&xsd;int”/>
</owl: DatatypeProperty >
<owl:DatatypeProperty rdf:ID=”hasFPG”>
  <rdfs:domain rdf:resource=”#user”/>
  <rdfs:range rdf:resource=”&xsd;int”/>
</owl: DatatypeProperty >
<user rdf:ID=”somchai”>
  <hasName rdf:datatype = ”&xsd;string”>Somchai</hasName>
  <hasAge rdf:datatype =”&xsd;int”>45</hasAge>
  <hasFPG rdf:datatype=”&xsd;int”>130</hasFPG>
</user>
```

Structural Knowledge

Instantiation

Figure 2-9: Diabetes knowledge modelling, Source: (Buranarach et al., 2009)

The same approach was used by Chalortham et al. (2009) to develop an ontology for Type II Diabetes.

2.3.1.1.3 Type II Diabetes Ontology

Chalortham et al. (2009) developed a type II diabetes ontology for clinical support system. This ontology tried to address the problems of interoperability within the healthcare system and offer systems and frameworks to healthcare providers that support and drive their processes.

The ontology building process used domain experts and diabetes clinical practice guidelines for knowledge acquisition and class hierarchy conceptualisation. Based on this ontology, a reminding system was developed that enables healthcare providers to suggest activities to improve the quality of life and manage diabetes treatment for patients. Figure 2-10 shows the architecture of this system.

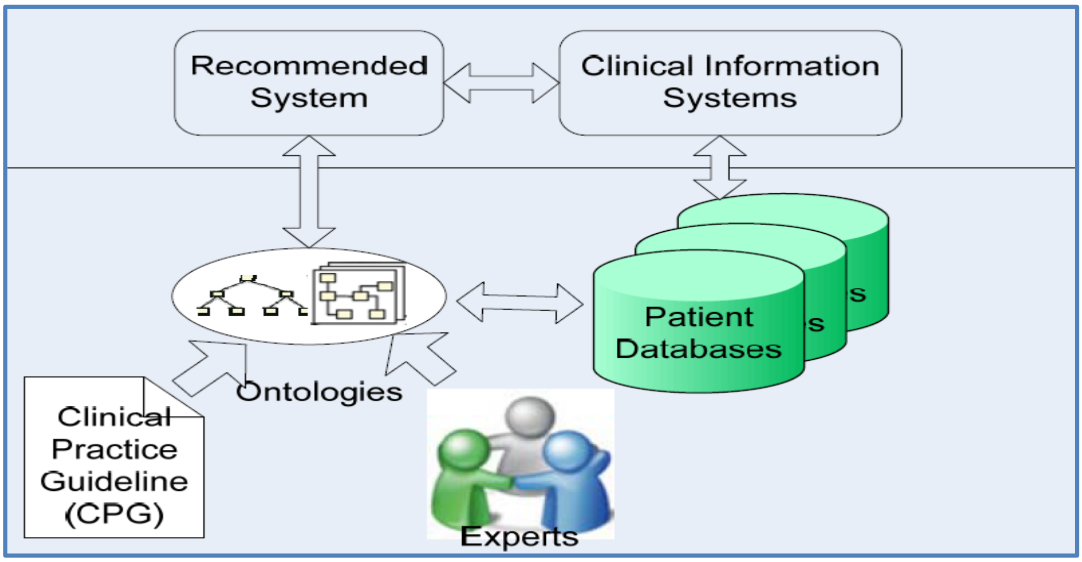


Figure 2-10: Architecture of ontology-based clinical support system, Source: (Chalortham et al., 2009)

It is observed that this ontology is costly and difficult to implement due to the reliance on domain experts, who are costly to employ and sometimes may not agree on what to include within the ontology, or what relationships exist between concepts in the ontology. This ontology reflects the expert terminology used by clinicians as it was developed based on the Thailand Diabetes Mellitus Clinical Practice Guideline. It thus lacks coverage of layman terms in this area and thus cannot be used for supporting analysis of forum discussions on diabetes.

Apart from using diabetes ontology in a knowledge management framework, other diabetes ontology have been developed for the study and the analysis of genetic susceptibility to diabetes.

2.3.1.1.4 Ontology for Glucose Metabolism Disorder (OGMD)

Lin and Sakamoto (2009) defined an Ontology for Glucose Metabolism Disorder (OGMD) in order to study genetic susceptibility to diabetes. Glucose Metabolism Disorder is a pathological condition where the glucose level in blood cannot stay within the *normal* range. OGMD included:

i) The names of the human disease, and

ii) Phenotypes and their observed parameters.

OGMD was used with the Ontology of Genetic Susceptibility Factor (OGSF) in order to define the susceptibility factors to diabetes. Figure 2-11 shows the top level of OGMD, which has been classified into diabetes complication, diabetes mellitus, hyper- and hypoglycaemia, and Pre-diabetes syndrome.

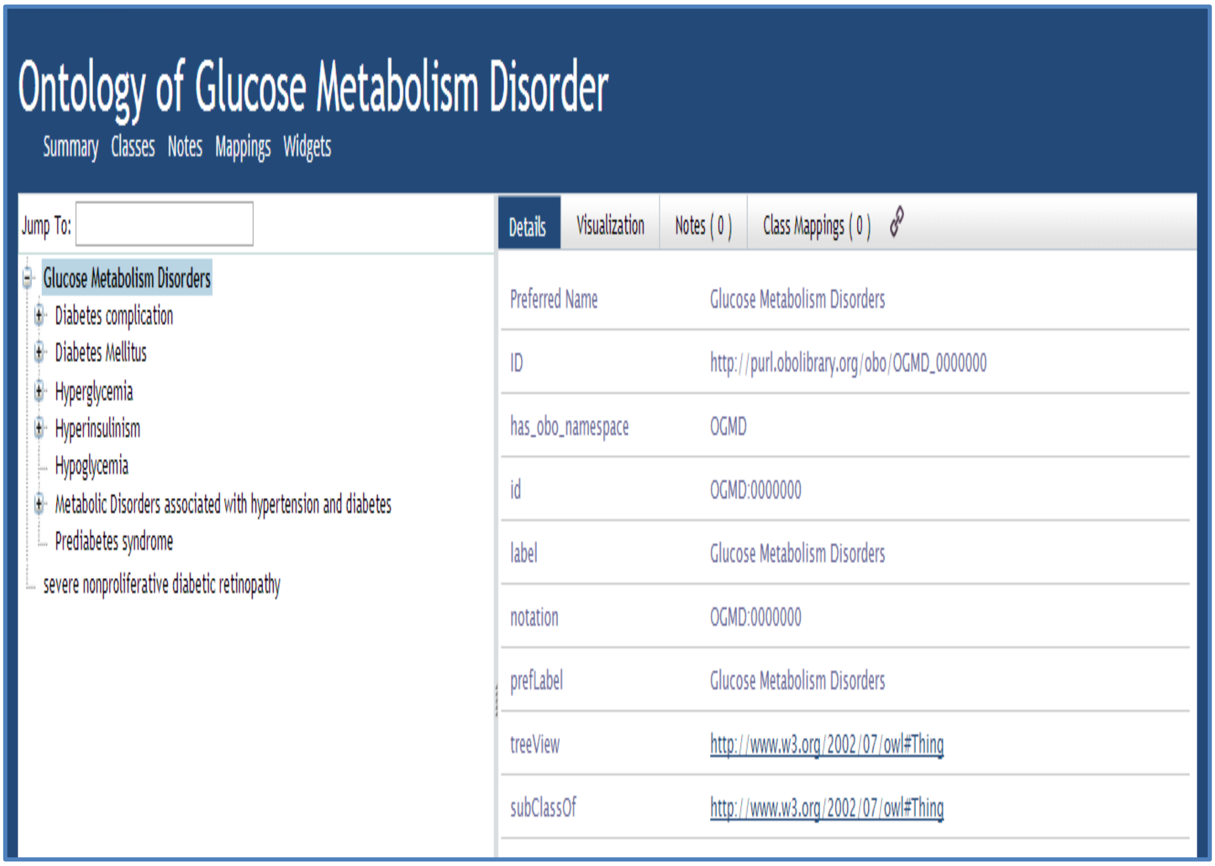


Figure 2-11: Ontology of glucose metabolism disorder, Source: (BioPortal, 2014)

From the viewpoint of the text analysis of diabetes forums, the main issue with OGMD is that it was developed for a specific application: “genetic susceptibility to diabetes”. This has guided the development of the ontology to include the concepts that are related to this topic and to ignore other concepts that might be related to diabetes in general. For example, this ontology does not contain any terms or concepts regarding “diet” or “medication”. In addition, the relations between the concepts are simple “parent-child” relations with no cross-relations at the same level. This limits the rich internal relations that are required to bridge the gaps between the concepts.

2.3.1.1.5 Summary of the previous work on Diabetes Ontologies

Different diabetes ontologies have been proposed in the literature to facilitate knowledge management applications and frameworks. The majority of these ontologies were built by domain experts (such as KBTA), or with a significant contribution from domain experts (such as OGMD). The application requirements and environment in these cases have guided the development of the diabetes ontology. Concepts were included if they were deemed suitable for the application while others were ignored. For example, some diabetes types were eliminated while developing the ontology as the application was concerned with specific types, such as in Type II Diabetes Ontology (Chalortham et al., 2009) and KBTA (Shahar et al., 1994). This can lead to poor information analysis and categorisation of unstructured information, and poor performance from the ontology-based system for the text mining of the Diabetes.co.uk forum.

Since the application of this research is in the broader medical domain of diabetes, it is worth exploring the rich literature and previous efforts in creating more general medical ontologies and terminology.

2.3.1.2 Medical Ontologies and terminology

Historically, the medical sector has been one of the first sectors to attract the attention of researchers for building ontologies and lexicons. They are commonly built to facilitate the sharing, analysis and reuse of medical information (Bodenreider and Burgun, 2005), such as patient records and diagnosis history. Various healthcare organisations have produced large structured ontology and lexicons. Key ones include UMLS, OpenGalen and SNOMED CT, which contain clearly defined concepts and their related descriptions as well as the relationships between them.

2.3.1.2.1 OpenGALEN

OpenGALEN is a project to develop an ontology based on a Common Reference (CORE), which is a model for clinical terms (Rector et al., 2003; Ivanović and Budimac, 2014). OpenGALEN ontology represents medical categories plus appropriate description about those categories in order to allow them to be categorised automatically. OpenGALEN proposed a hierarchy structure that comprised four layers (Figure 2-12):

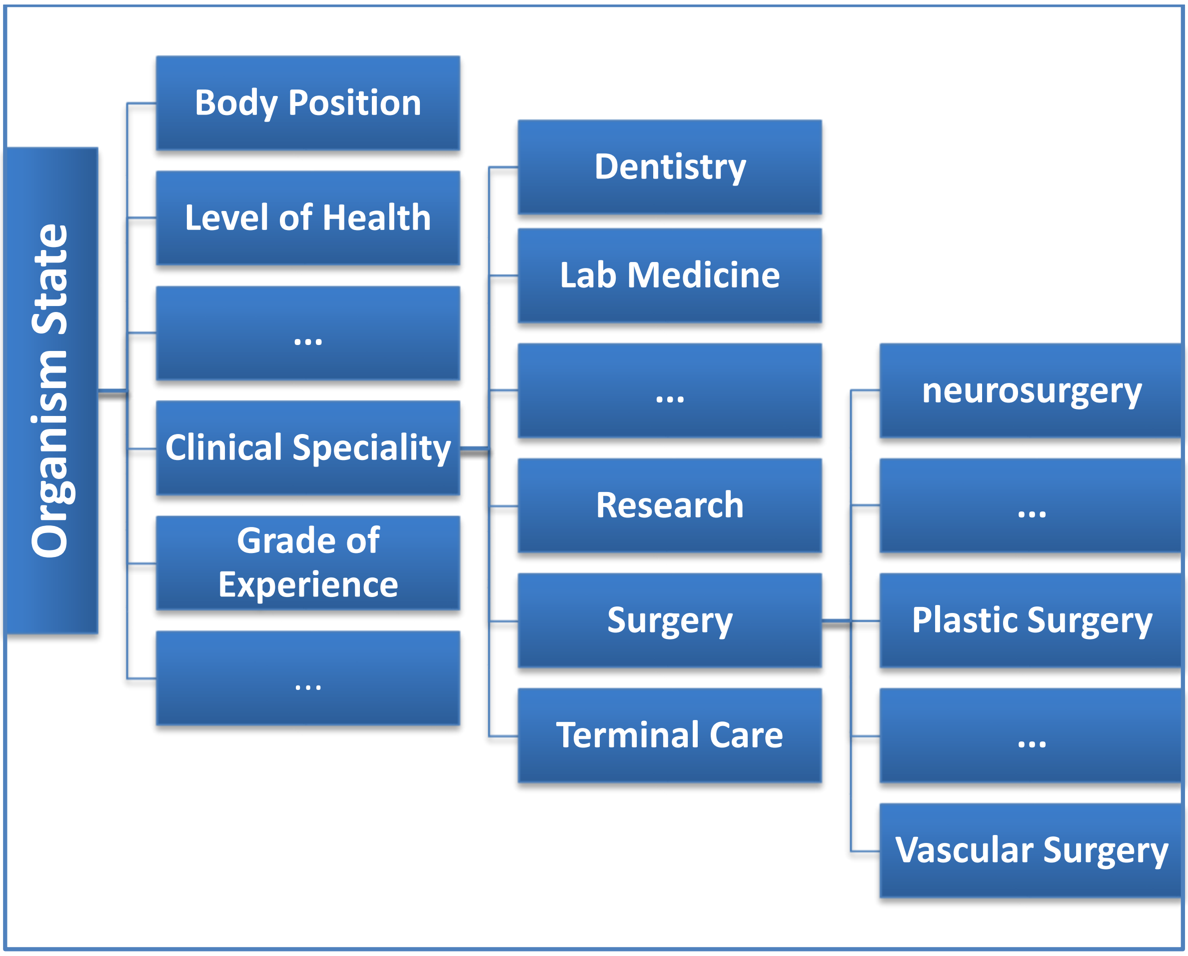


Figure 2-12: OpenGALEN hierarchy

1- High-level Ontology: This part represents and describes the major categories of the structure (Such as OrganismState, ProcessState, and StructuralState);

2- The Common Reference Model (CRM): This part represents the reusable part of OpenGALEN and it is application-independent model of the medical and clinical terms and their definitions/descriptions (Such as AgeState, AnaesthesiaState, and ClinicalSpecialityState).
3- Building Block's "Detailed Extensions", which are required for describing specific sub-domains/sub-categories, such as Dentistry, Research, Surgery and TerminalCare.
4- Surgical procedures model and related models to define and describe complex and combined terms, such as Faciomaxillary Surgery, Neurosurgery, Orthopaedics and Vascular Surgery.

It was observed that OpenGALEN ontology did not provide sufficient concepts in the medical domain: The ontology often excludes the dictionary equivalent of the main terms. For instance, "Plastic Surgery" is included in OpenGALEN as a sub-category of "Surgery", but "Cosmetic Surgery" is not recognised as a concept/description in the terminology.

Another issue with OpenGALEN is that even though its CRM model defines clear hierarchical layers, there is no clear dividing line between these layers, since they are defined organisationally instead of having firm logical conditions to separate them. As a result, there are no clear boundaries between the layers (Rogers, 2004). Nonetheless, OpenGALEN ontology employed this hierarchy because of its benefits when used in automated applications.

In comparison, SNOMED-CT has tried to offer the required rich core relationships for its terminology.

2.3.1.2.2 Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT)

Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) (IHTSDO, 2014) development was started by the College of American Pathologists as early as the 1960s and was known as SNOP at that time. It was re-launched in 1999 through the merging and expansion of the SNOMED Reference Terminology (SNOMED RT) and the Clinical Terms V3 (CTV3) reference, which was developed by the UK National Health Service (NHS).

SNOMED contains 18 autonomous hierarchies, which echo the structure of earlier versions of the terminology and organise them into several "blocks", such as "Living organism", and "Procedures".

In order to build rich internal relations within SNOMED CT, the team consulted with many medical and clinical professionals for the alpha test from six countries and as a result identified over three hundred thousand concepts and over a million relationships between them (IHTSDO, 2014). The types of these relationships are different to those presented in OpenGALEN since concepts are categorised from different points of view through different connections/descriptions with other concepts. The rich "network of relationships" within SNOMED greatly improved the internal links and offered many connections to other domains/fields.

Compared to OpenGALEN, SNOMED populated the vague boundaries in OpenGALEN by filling the gaps between the terms using shared descriptive concepts. In this organisation of the concepts, two terms can mutually explain each other and thus address the shortage of internal relations in OpenGALEN. For example, SNOMED tolerates that a child concept might have multiple links to different parent nodes in order to enrich both the vertical and horizontal connections within the structure.

Theoretically, SNOMED might be the ideal medical lexicon to be utilised in this research application. However, SNOMED has three main weaknesses in the context of mining diabetes forum text data:

i. **Size of the terminology**: SNOMED has more than 300,000 unique concepts and over one million internal relationships. However, several researchers argue that it has become too large to be effectively and efficiently reused/sustained (López-García et al., 2012). SNOMED CT requires highly skilled experts in order to be applied/used in different applications and its size and complexity requires significant investment in IT processing capability. Using a customised version of SNOMED CT might be an alternate approach, and is recommended by IHTSDO for small applications.

ii. **Customisation**: SNOMED CT depends on the organisations that want to reuse it, to develop their own customised segments that are appropriate/suitable for their applications. Several segmentation approaches have been developed that are automated and are able to limit the need for a big terminology. For instance, Noy and Musen (2001) argue that a subset could be generated through defining concepts and relationships that represent the segment and then mining all the concepts and relations connected to them. Another approach to highlight the subset might be by only traversing the hierarchy and adding the concepts and relations in all definitions of the atomic terms in the segment vocabulary, without taking into consideration the other dependencies (López-García et al., 2012). These approaches proposed traversing the structure for a specific target application. However, when a large terminology such as SNOMED is the source for the target ontology, the required computational efficiency is a big barrier for building and forming sub-ontology for the specific needs of the applications (Bakhshi-Raiez et al., 2012). In addition, these techniques are intended to derive parts of the ontology from a domain viewpoint and not from an application viewpoint. Using SNOMED CT for this research would need a re-building of a diabetes sub-section using the concepts and relations from SNOMED CT. In order to build a new subset/segment, the functional requirements of the application plays a significant role in arranging application-related terms and the connections between them. Customising the structure of the ontology to suit the application viewpoint has been highlighted as a fundamental research problem by authors such as Jarrar and Meersman (2009).

iii. **Coverage of Content**: The main focus of SNOMED CT development has been to facilitate working between medical professionals, such as in information sharing. It has not been designed to improve the categorisation of natural language medical data. Even though SNOMED CT contains a vast number of concepts and relations, they still do not extend sufficiently to cover the more general concepts and relations. Some differences stem from the absence of a number of terms in SNOMED CT that other standards or ontology contain (Nachimuthu and Lau, 2007). Several studies have shown that SNOMED CT has a range of coverage of between thirty and ninety percent for different clinical domains (Brown et al., 2005; van der Kooij et al., 2005; Kim et al., 2005; Lee et al., 2013; Liu et al., 2012).

Since SNOMED CT was built by medical professionals for the profession, it has a good coverage of the viewpoint of medical specialists. However, in the context of this research, it does not represent the terms and concepts used by non-experts sufficiently. For example, patients use the term “tight control” to describe keeping the glucose level as close to normal levels as possible. SNOMED CT does not contain such a term, instead it contains “Blood glucose control” and “Blood glucose control education” as the closest terms to “tight control”. This point is supported by authors such as Schulz et al. (2009) who have criticised SNOMED CT for not having a sufficient number of relations and connections with non-professional concepts.

The Unified Medical Language System (UMLS) project argued that the collection of medical information directly from the medical professionals was a main cause behind these issues in SNOMED CT. Hence, it favoured exploiting and utilising the published resources and existing ontologies in order to construct the UMLS ontology.

2.3.1.2.3 The Unified Medical Language System (UMLS)

The Unified Medical Language System (UMLS) project was created by the United States National Library of Medicine (NLM) in order to build a database that integrates and distributes key biomedical ontology/terminology, as well as coding standards and associated medical sources. It utilises several medical ontologies and terminology such as Medical Subject Headings Ontology (MeSH®) (NLM, 2014), SNOMED CT® (IHTSDO, 2014) and Logical Observation Identifiers Names and Codes (LOINC®) (Anna, 2010). These reused sources are the bases for building the large, multi-purpose terminology database in UMLs, which is called “Metathesaurus”. This Metathesaurus is organised by terms and meanings, and is linked to the other tools provided by UMLS: the “Semantic Network” and the “SPECIALIST Lexicon and Lexical Tools”. The semantic network tool consists of the “semantic types” (i.e. categories) and their connections (i.e. semantic relationships). It includes and integrates multiple resources via linking their terms utilising their semantic connections, while the “SPECIALIST Lexicon and Lexical Tools” are tools provided by UMLS for applying the natural language processing of the terms and codes (NLM, 2013).

The Metathesaurus within UMLS is very large and complex since it is a compilation of about a hundred individual medical vocabularies. It causes significant comprehension issues for the users and it is difficult to integrate new terminology into the existing structure (Gu et al., 2012).

Although the exploitation of existing knowledge resources speeds up the creation process of new ontology and results in lower costs, these sources are complex and their update process can be very slow compared to the information collection from specialist and non-specialist users. For instance, a new virus named Middle East respiratory syndrome coronavirus (MERS-CoV) was discovered and reported in the medical journals in 2012 and was only included in SNOMED and UMLS two years later. Another example is a popular diet amongst diabetics, which is the Low Carbohydrate High Fat (LCHF) diet. When searching SNOMED and UMLS, it turns out the term was not yet included in any of them by April 2014. There is a need in various healthcare applications to obtain the latest medical information. In fact, within social media analytics applications for healthcare purposes, there is a challenge to acquire up-to-date healthcare trends that users are sharing, and it is a difficult task for these applications to capture the changes quickly and in a complete manner.

2.3.1.2.4 Discussion of Medical Ontology and Terminology

The review of the needs and requirements for the research case study, as well as the review of several medical ontologies have highlighted the practical problems in creating or reusing them for this research. Nonetheless, the discussion has revealed several features that a suitable ontology should include:

- **Width and depth of internal relationships**: Previous work has suggested the use of rich relationships in order to improve reasoning. An ontology that targets wide domains should include a network structure similar to SNOMED CT. The ontology relations for a specific domain should connect the concepts from various viewpoints (multifaceted ontology).

- **Semantic Relationship Utilisation**: Medical ontologies using linguistic semantic relations, such as UMLS, have shown the capability of providing relations between professional and non-professional terminology.
- **Reuse of Knowledge Sources**: It has been demonstrated by various medical ontologies that the reuse of the accessible knowledge sources may help in minimising the dependency on domain experts, and accelerate the process of ontology engineering. The reuse of such sources could become a good alternative for quickly building a particular ontology for this research.
- **Speed of Ontology Updating Process**: The ability of various applications to address the fast knowledge changes in the domain remains a significant issue when developing ontologies. Current methods in different domains seem to be poor when collecting and integrating new concepts within a domain ontology. These difficulties stem from the cost of using domain experts to get direct ontology updates, or from the difficulty of finding and incorporating indirect sources which offers that ability. This problem is even more significant in cross or multi-domain ontologies. It could be argued that with the increasing specialisation of work, but increasing value from multi-disciplinary collaboration, this is a major problem.
- **Lightweight vs. heavyweight**: A “Lightweight ontology” is one that is reasonably flexible when defining its concepts and the connections between them, while a “heavyweight ontology” is a lightweight ontology with added constraints to strictly define the concepts and the connections between them (Uschold and

Gruninger, 1996). For example, an ontology could have a relation that connects "nitrogen mustard" and "Cancer" through $\xrightarrow{\textbf{\textit{is related to}}}$ relationship, which is considered to be a very general semantic relation. However, if another ontology contains both "$Cancer\ treatment \xrightarrow{\textbf{\textit{is a property of}}} Cancer$" and "$nitrogen\ mustard \xrightarrow{\textbf{\textit{is a type of}}} Cancer\ Treatment$" relationships, which are used to define concepts and their descriptions, the second ontology can be considered a heavyweight ontology, while the first one can be seen as a lightweight one.

In this research, a supporting ontology is considered essential to act as a bridge between specialist and non-specialist terminology. In this ontology, concepts describe each other. There is a mixed use of terms and descriptions, while semantic relations facilitate the coverage of non-specialist area. The use of axioms to explicitly represent connections is NOT forced. These characteristics of lightweight ontology show that it is possible to have flexible concepts and connections' definition, when a rich structure is available.

The review of medical ontologies, terminology, and diabetes ontologies in particular highlight the need for linguistic relations between the concepts, which can be obtained using linguistic ontologies that focuses on the connections between terms rather than on a specific domain.

2.3.1.3 Linguistic Ontologies

The purpose of building a linguistic ontology is to model natural languages through capturing their concepts and terms, as well as their connections within grammar elements in order to support constructing other ontology (Gómez-Pérez et al., 2004). A few linguistic ontologies have been proposed to offer concepts and their relations from a linguistic perspective. Key examples of these approaches include WordNet, SENSUS, and Microkosmos.

2.3.1.3.1 WordNet

WordNet is an example of ontology with a broad coverage. It is a large online lexical database, created at Princeton University in order to model the English language domain. It was structured based on unique lexical concepts (known as Synsets)[5] rather than their formation (PrincetonUniversity, 2010). In addition, WordNet provides relationship information via its complex concept hierarchy (Bengoetxea et al., 2014).

WordNet is used as a knowledge base of concepts, their types and their relationships in order to offer a broad, yet beneficial, background knowledge in the domain of English language. A synset shows a unique occurrence of a term meaning, which is related to other synsets using a specific relationship. These natural concept relationships were collected via focusing on the more sophisticated lexical connections. Lexical relationships in WordNet uncover human language relations in a constraint sense that do not include the full semantic relation. For instance, when searching for “grey” in WordNet (Figure 2-13), a relation to “any organization or party whose uniforms or badges are grey” can be found. However, it is hard to link "The Greys” political party in Germany with “grey” as there is no lexical relation between the two in WordNet.

[5] WordNet V3.0 contains approximately 117,000 unique Synsets

Figure 2-13: WordNet entry for the word "Grey"

The hierarchy structure of WordNet was enhanced using a network of vocabulary relationships. These relationships looked at the concepts from different viewpoints and recognised them in various ways, such as by using different paths to other concepts.

This deep inner structure offers sufficient information for WordNet to include any new concepts or terms and incorporate the new information as a part of WordNet. However, the WordNet internal structure is language orientated without any specific domain requirements. Consequently, any attempt to extract an appropriate part of WordNet to suit a specific domain is very complex, and may require going through all the terms to identify the possible domain concepts as well as recalculating the relationships between the domain related concepts. Since WordNet was largely designed to operate as a background database for various natural language processing (NLP) and text mining applications, it difficult to apply in any domain not covered completely by the terminology in WordNet. Similar problems occur in other general ontologies that do not cover a specific domain.

To address this issue, other approaches, such as the Mikrokosmos methodology, have been proposed.

2.3.1.3.2 Mikrokosmos

The Mikrokosmos ontology methodology (Renamed as OntoSem) focused more on the explanation of relationships between the concepts/terminology in the language, especially with regard to their settings and situation, in addition to their “natural meaning” within the language interpretation (Nirenburg and Raskin, 2004). In practice, Mikrokosmos highlights the distinction between experts’ concepts and general concepts, as well as their occurrences in the language ontology. These modification in Mikrokosmos attempt to fill the gap between linguistic and domain ontologies. Mikrokosmos showed how linguistic relations could be utilised in domain ontologies in order to translate terms between experts and non-experts, or for synthesising multi-domain ontology. A similar approach was also utilised in the SENSUS ontology.

2.3.1.3.3 SENSUS

The SENSUS ontology methodology was developed by the Natural Language Group (NLG) at the Information Sciences Institute (ISI) in the University of Southern California with an emphasis on natural languages (Swartout et al., 1997; Iqbal et al., 2013). It was built to enhance different applications such as text summarisation, information retrieval and machine translation tasks through the identification and interpretation of meaningful semantic relations between concepts.

The top-level section of SENSUS ontology was constructed to include the essential terms/concepts and their relationships from existing knowledge bases and dictionaries, such as WordNet and the Collins dictionary (Figure 2-14).

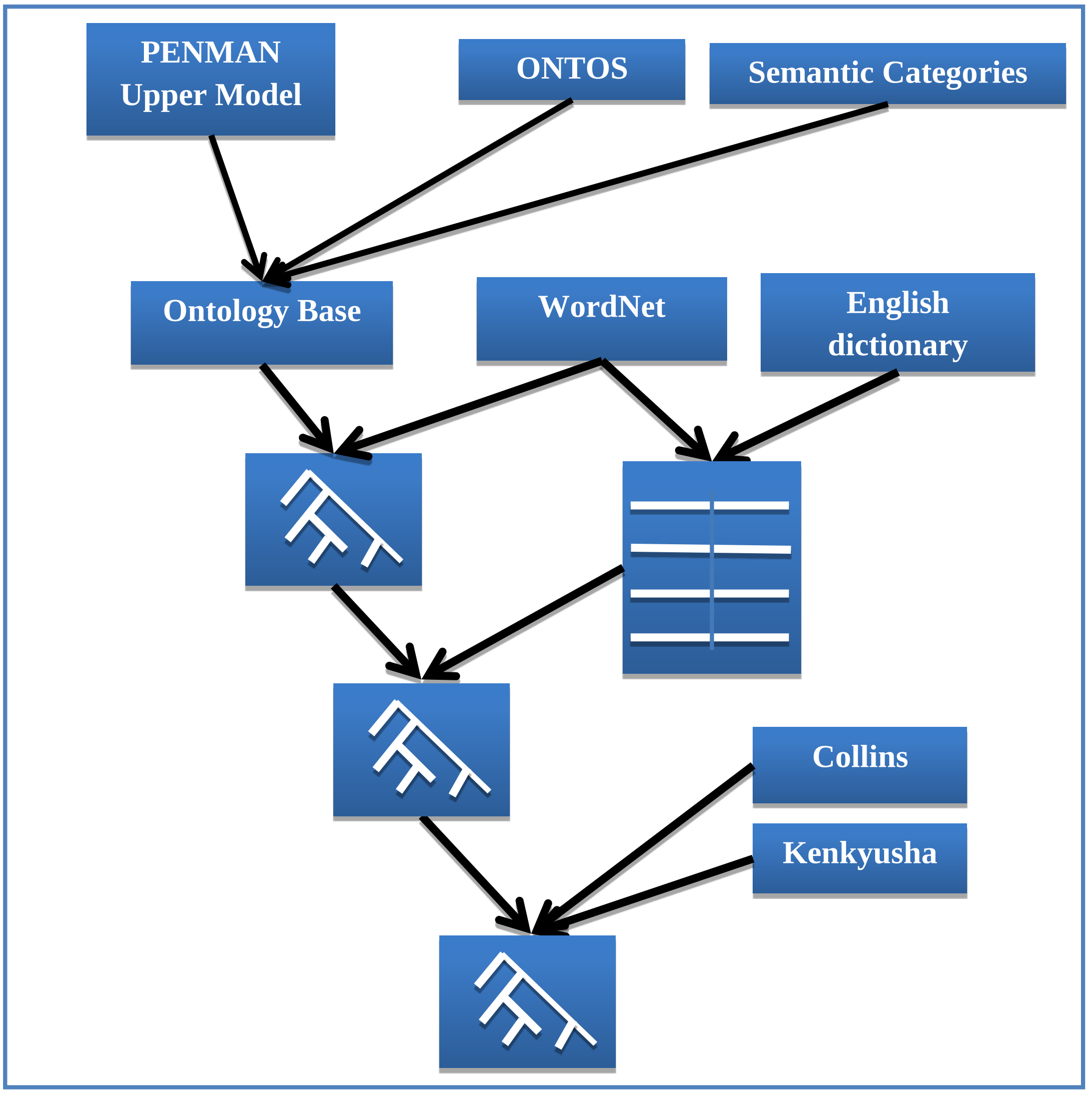


Figure 2-14: SENSUS merging strategy, Source: (Swartout et al., 1996)

The SENSUS methodology builds an ontology for a specific domain based on the already established knowledge bases, or preferably, an earlier big ontology. However, the SENSUS methodology does not employ the traditional reuse process: It detects the key concepts for the domain (called "seeding words") and connects them to the existing big ontology/knowledgebase. Then, unrelated concepts to the new domain can be "trimmed" from the big knowledgebase (Iqbal et al., 2013). Figure 2-15 shows the process that SENSUS methodology undertakes when producing a new ontology:

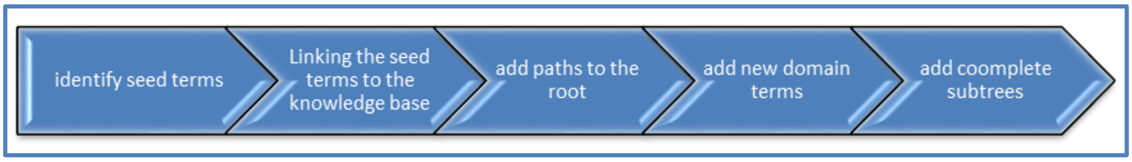


Figure 2-15: SENSUS methodology for ontology engineering, Source: (Ma et al., 2014)

One distinguishing feature of SENSUS methodology for constructing ontology is the way by which semantic relationships are recognised. This method was established based on the hypothesis that any definition or description of a specific concept has a few, yet highly related, group of concepts that enable the semantic relationships between this concept and other concepts in the ontology.

For example, *Happiness* can be described by a group of other words: *joy*, *rejoice*, *felicity*, *pleasure* and *delight*. This reveals wider semantic relations within the ontology: A specific term/word can be represented or defined by several terms that are related to it. This approach shows that ontologies may use a different and broader type of relationship compared to the lexical relations found in WordNet: A co-existing relation between concepts is derived from the domain.

Another unique feature of SENSUS methodology is the way it changed the structure of the linguistic ontology from a simple tree structure to a more complex network structure. A network structure can increase the number of internal relations vastly and, as a result, more advanced reasoning tools can be utilised with the ontology. In addition, the network structure can help investigate the ontology from different points of view (facets). This means that the ontology would be more flexible when interacting with different domains, or would be easier to structure into various formats depending on the needs of different applications.

Nevertheless, both ontologies (SENSUS and Mikrokosmos) utilise previous linguistic ontologies (such as WordNet) and as a result, they may suffer from the issues discussed in section 2.3.1.3.1.

2.3.1.3.4 Discussion of Linguistic Ontologies

The use of semantic relations, especially the "co-occurrence" type suggested by SENSUS, may be very useful for this research: Firstly, this type of relationships resulting from the co-occurrence, could have the potential to overcome the restrictions in lexical relationships imposed by linguistic ontology. Consequently, more complex and quantifiable relations can be generated. This research could apply these types of relationships in order to categorise concepts that are semantically related.

Secondly, the network structure of the ontology and the coverage of the domain from different viewpoints can bridge the gap between concepts used by specialists and non-specialists. This is very useful when analysing online discussion forums, since the users of these forums employ both types of terminology when discussing domain-specific topics.

Semantic relationships that are required in building a diabetes ontology for analysing the diabetes discussion forums may be created with the support of a linguistic ontology in order to enrich the structure and incorporate non-specialist concepts and terminology.

2.3.1.4 SEA

The "Semantic (S) relatedness oriented ontology engineering via retrieving information from the search Engine (E) index with assistance from social network Analysis (A)" (SEA) is an approach that has been proposed to address the issues of ontology engineering using concepts and semantic relationship mined from the Google Search Engine Index. It utilises social network analysis techniques to carry out the ontological structure analysis. This approach was tested in both engineering and medical sectors to produce domain ontologies that provide breadth and depth of domain coverage through forming a network structure that allows concept investigation to be carried out from multiple points of view (Ma et al., 2014).

The output of SEA is a multi-faceted ontology in the form of a network, and it addresses several challenges regarding

- The cost of generating the ontology, through less reliance on domain experts;
- Scope and relationships richness;
- Breadth and depth of concept coverage.

The evaluation of the SEA ontology conducted by Ma et al. (2014) for the engineering sector case study showed that it had a better performance when substituted in automated information applications that used traditional ontologies (the case study was an ontology-driven marketplace for engineering businesses, WMCCM[6]). The performance improvement was realised through the rich semantic relationships and a broader domain coverage (Figure 2-16).

[6] www.wmccm.co.uk

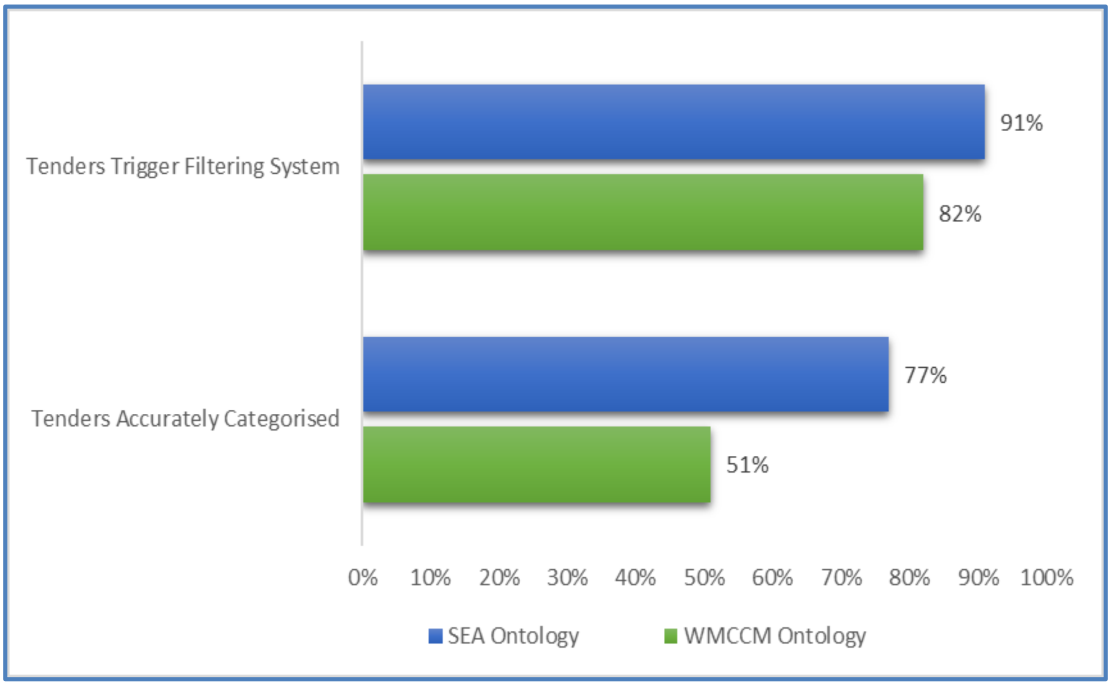


Figure 2-16: SEA vs. traditional manufacturing ontology, Source: (Ma et al., 2014)

This relatively new approach has unique features that provides advantages compared with the previous approaches discussed:

I. A main advantage of this approach is that SEA does not require domain experts to provide the overall ontology concepts and relations. Instead, it takes six initial *seeding keywords* and then automatically links them to the knowledgebase. This technique reduces the cost of building the ontology as it constrains and speeds the process of collecting information.

II. An important characteristic of SEA is the multifaceted ontology output. The derived ontology is a network of concepts that can viewed from multiple perspectives, which allows the ontology to be easily customised to suit different applications. SEA uses Social Network Analysis (SNA) in order to clarify the network structure of the ontology.

III. The relationships between the generated concepts are weighted and directed, which enhances the logical reasoning.

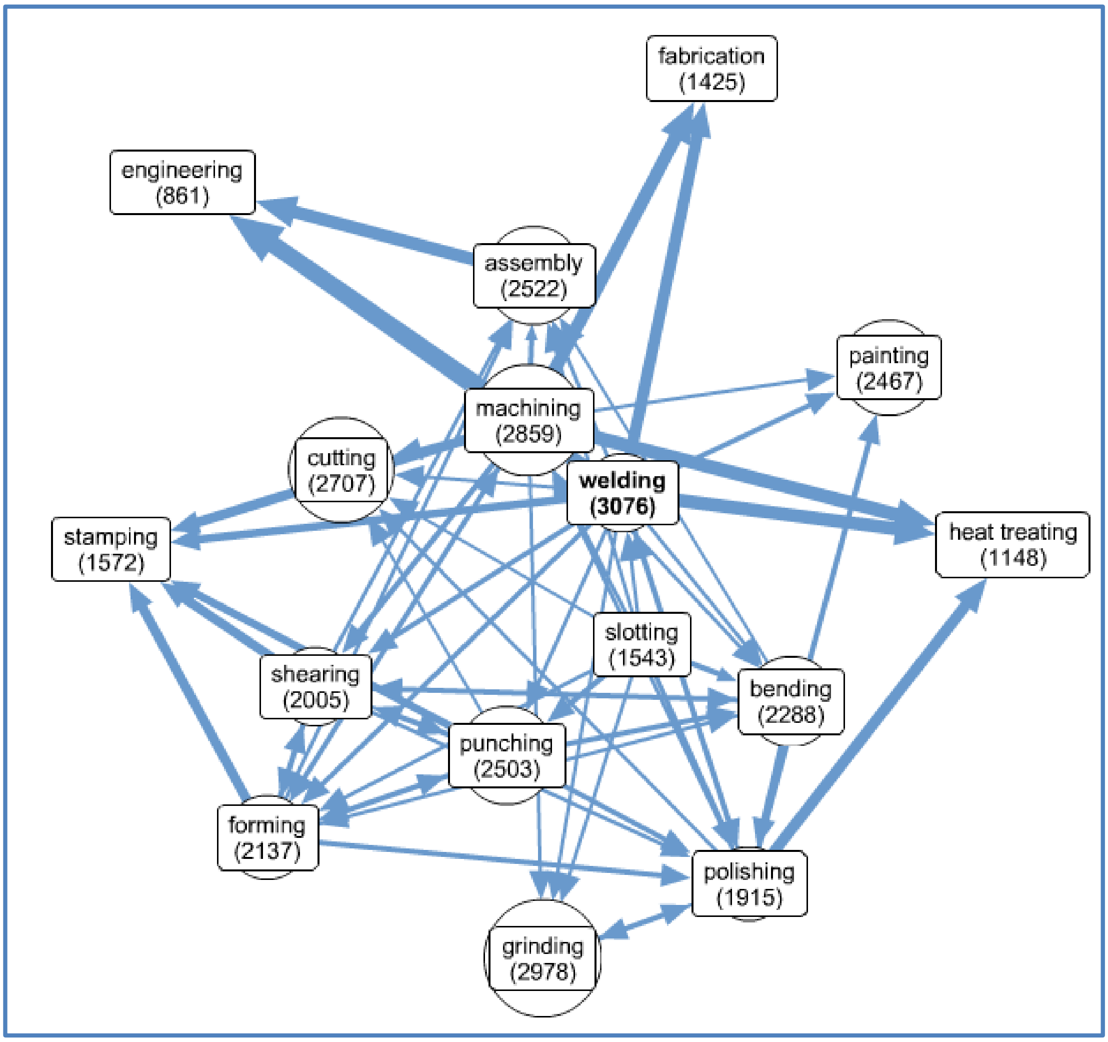


Figure 2-17: SEA Engineering Ontology

This approach seems more aligned with the purpose of this research. However, SEA was developed for producing ontologies suitable for resource matching applications rather than analysing social media text, where text enrichment, text annotation and disambiguation are the main tasks for the text analysis. SEA claims that it links the generated terms and concepts to the Google Index knowledge base, yet this linkage is implicit and cannot be tracked to the source. This means that it is very difficult to perform tasks such as text enrichment using the derived ontology. In addition, SEA utilised the "GoogleSets" tool in order to link seeding keywords to the Google index and identify semantic relations. However, GoogleSets as a tool is no longer available.

It may be valuable to explore and identify a new source of concepts that can enhance the output of a SEA type, seeding word driven tool, to produce an ontology suitable for text mining. The tool must produce semantic related concepts using a limited number of seeding words quickly and reliably. The ontology corpus construction process discussed in chapter three describes approaches to achieve these attributes.

The preceding discussions have highlighted the value of expert and novice domain knowledge to the analysis of domain specific textual data as in the Diabetes.co.uk forum.

2.4 Introducing Domain Knowledge into Text Mining Systems

Domain knowledge can be utilised in text mining applications in several ways and at different phases within the architecture of a text mining framework. There may be many ways that domain knowledge can enhance the outcome of knowledge discovery on a given text corpus. Three key practical arguments for including domain knowledge in text mining are (Feldman and Sanger, 2007):

- Domain knowledge can be used in a text pre-processing phase in order to produce a reliable and constant reference area, as well as a suitable hierarchy for terms and concepts, which will be very beneficial later through other phases, such as representation and refinement.
- Domain knowledge can be utilised in order to limit/restrain the number of possible connection patterns: Domain knowledge can create constraints that help build more efficient and focussed queries. These constraints may also be used for other objectives and in other phases of text mining;
- Domain knowledge can provide a very efficient method for solving problems such as the synonymy and polysemy of terms and concepts at different levels

of the architecture. For example, the availability of an ontology that contains both lexical and semantic relationships can help utilise different types of resolution mechanisms.

Possibly, the easiest mechanism to integrate domain knowledge within a text mining framework is through utilising it in the building of meaningful constraints. For example, the terms in the text mining framework can be sorted and categorised into structured hierarchical classes or a cluster (Feldman and Sanger, 2007). These classes can later be matched against a relevant external ontology in order to obtain beneficial features for these classes and the relationships between them.

Domain knowledge-based constraints may also be utilised in a novel way in the presentation phase through the formatting of text mining results. A practical example for this could be while displaying the results of a specific query for all the enterprises and countries associated with "*wheat*", the text mining system might highlight the companies and countries that **supply** "wheat" in a colour (using green for example) different from the colour associated with companies or countries that **buy** "wheat" (using red for example). This use of colour coding (which is enabled through an external knowledge source) help user's in evaluating the data, because this coding offers more and easier-to-interpret information than to simply showing a plain list of associations to the user separated only by the level of confidence in the results.

Another use of background knowledge in text mining is in the production of hierarchies of concepts in the text corpora. During the pre-processing phase of the text mining system, sets of terms may be assessed against a hierarchical concept structure produced from the background knowledge (ontology). The output hierarchical structure of concepts contains the advantage of being influenced by the ontology and its relations as well as being well-incorporated within the external source (Feldman and Sanger, 2007). This can be useful when other system processes require access to concepts and relations contained in the ontology. In social media, domain knowledge has been used primarily for semantically annotating text.

2.4.1 Semantic Annotation of Social Media

Semantic annotation is the process of binding semantics and language (usually in a textual form) together. It could be described as the formation of bidirectional connections between background knowledge (mainly in the form of an ontology) and unstructured text (in the form of documents or social media posts) (Kiryakov et al., 2004). From a technical point of view, annotation is about tagging all the terms and concepts mentioned in the text from the background knowledge, usually through referencing the Uniform Resource Identifiers (URIs) of these concepts in the background knowledge.

Semantically annotating content produced by online users enables various processes and systems such as semantic search and browsing, semantic-based recommendation systems, semantic visualisation and analytics, as well as facilitating the building of semantic-based user behavioural models (Bontcheva and Rout, 2014). Semantic annotation is used in different contexts in fields such as electronic government, knowledge management and electronic health.

Semantic annotation can be done either manually via human annotators, semi-automatically via suggesting annotations to users who then edit and correct the annotations, or automatically (Bontcheva and Cunningham, 2011). An example of manual semantic annotation in the context of social media is the Semantic

Microblogging framework (SMOB) (Passant et al., 2010). SMOB allows online users to add semantics manually to posts and messages.

Hepp (2010) has proposed another manual *syntax* for semantic annotation for microblogging posts. These annotations are transformed into Resource Description Framework (RDF) data models. This syntax provides relationships between tags in microblogging messages such as sub-Tag or same-As, and supports the use of multiple statements in a single microblogging post.

While the manual approaches for semantic annotation are useful, there is a need to focus on automatic methods for semantic annotation in order to enable the analysis of the vast amount of posts on social medial websites. As a result, this research will focus mainly on the automatic semantic annotation.

One form of language analysis known as information extraction (IE) is playing a key role in inter-connecting the unstructured text data and background knowledge. One particular type of information extraction, known as "*ontology-based information extraction*" (OBIE) is particularly adjusted for the process of semantically annotating unstructured text using ontology (Li and Bontcheva, 2007). As its name implies, this type of IE differs from other types by using ontology as a main input in order to facilitate the task of information extraction, as well as having the ontology as an output of the information extraction process. If the output of the system is specified with regard to an ontology, without modifying or updating the ontology itself, then the method is termed *ontology-oriented information extraction (OOIE)* (Bontcheva and Rout, 2014).

OBIE and OOIE have another distinctive feature compared with other IE methods: they identify the extracted entities via connecting them to their semantic description in the background knowledgebase using URIs. This enables the tracking of entities across different messages or documents as well as enriching the entities description using the background knowledgebase. Practically, this needs processes and systems for automatically recognising entities, terms and the relationships between them. These are usually part of the second phase of text mining (text representation) and are usually preceded by linguistic pre-processing (discussed in section 2.1).

2.4.2 Ontology-Based Entity Recognition in Social Media

Entity recognition based on using ontology usually contains two main steps:

- Candidate Selection, and
- Linking.

Candidate selection (commonly referred to as entity annotation) focuses on identifying all the occurrence of the instances from the used ontology in the text. For example, using DBpedia[7] as an ontology (an online knowledgebase created by automatically extracting structured information from Wikipedia), *Mercury* can be an element, a planet, or a Roman God in mythology. The second step, i.e. entity linking (known as entity disambiguation), then utilises relations from the ontology beside the surrounding text to identify the context in order to disambiguate the entity and determine the best URI to link the entity to. It should be stated that not all entity recognition algorithms perform both phases, i.e. some methods just carry out the first step of identifying the occurrence of the entity in the text (Li and Bontcheva, 2007).

[7] http://dbpedia.org

2.4.2.1 Commercial Semantic annotation Services

Several commercial services exist to annotate documents (including web documents) and link them to different knowledge bases using URIs. This section focuses on two main services: Zemanta and Open Calais that have been either used or tested in the published literature, such as (Saif et al., 2012; Rowe and Stankovic, 2012; Abel et al., 2011).

Zemanta[8] is an online annotation commercial tool, which was developed initially to help users annotate and insert tags for blogs and emails using tag recommendations. Figure 2-18 shows an example of annotating text and the tags recommended by Zemanta, in addition to the potential links to Wikipedia and other relevant sources. Then it is the users who decide which tags to be used and applied and which URIs that they want to add. In Figure 2-18, the links in the text were created to highlight entities that point to the Wikipedia pages on the corresponding topics.

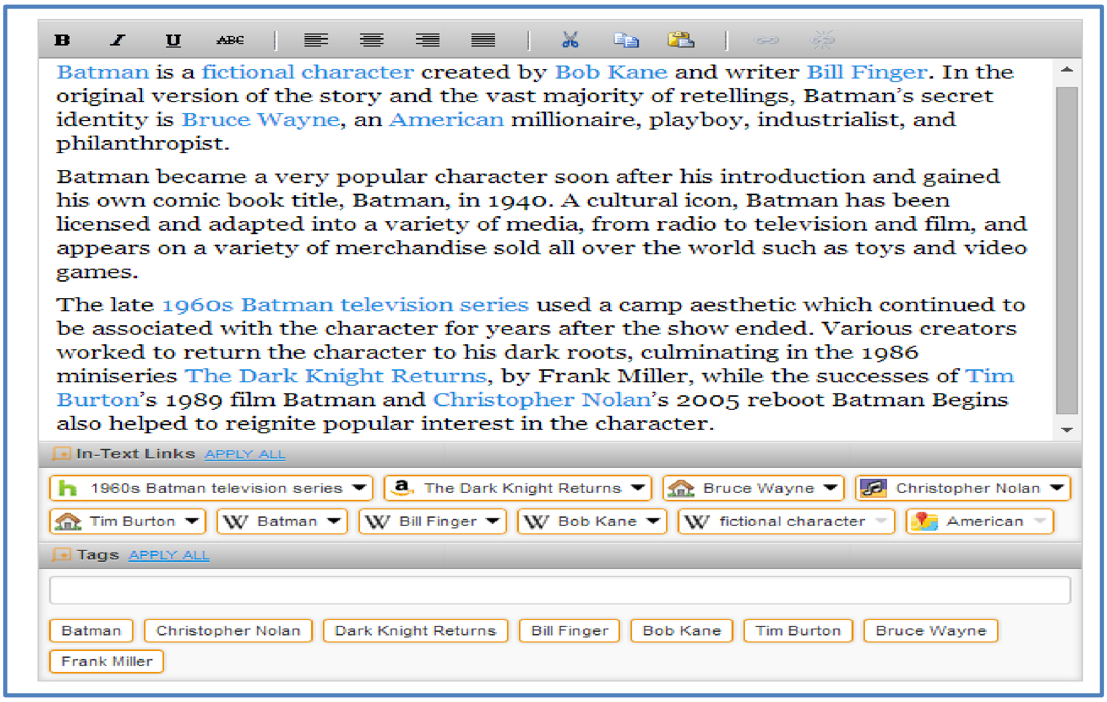


Figure 2-18: An example of Zemanta API

[8] www.zemanta.com

OpenCalais[9] is also a commercial online tool for entity annotation, which has been used by several studies to annotate text extracted from social media. For example, Abbar et al. (2013) utilised OpenCalais in order to annotate entities in the comments on the news articles. Then, the output was used to recommend other articles in real-time to the readers based on the entities and sentiment extracted from the article and the resulting comments.

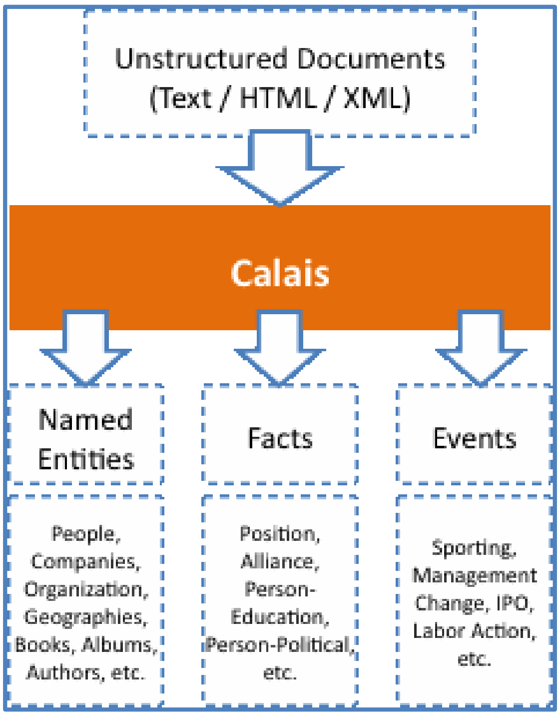


Figure 2-19: OpenCalais service, Source:(OpenCalais, 2013)

The extracted entities are typically people, companies and locations. In addition to entities, events that involve these entities are extracted, for example merger and acquisition, in the case of companies. Figure 2-20 shows an OpenCalais Viewer with an example of annotated text. The annotated text includes URIs that allow access to get more information about the entity using linked data. OpenCalais currently utilises eight Linked Data assets: DBpedia, Wikipedia, Freebase, Reuters.com, GeoNames,

[9] www.opencalais.com

Shopping.com, IMDB, and LinkedMDB. These resources generally relate to the entity types covered by the OpenCalais Ontology.

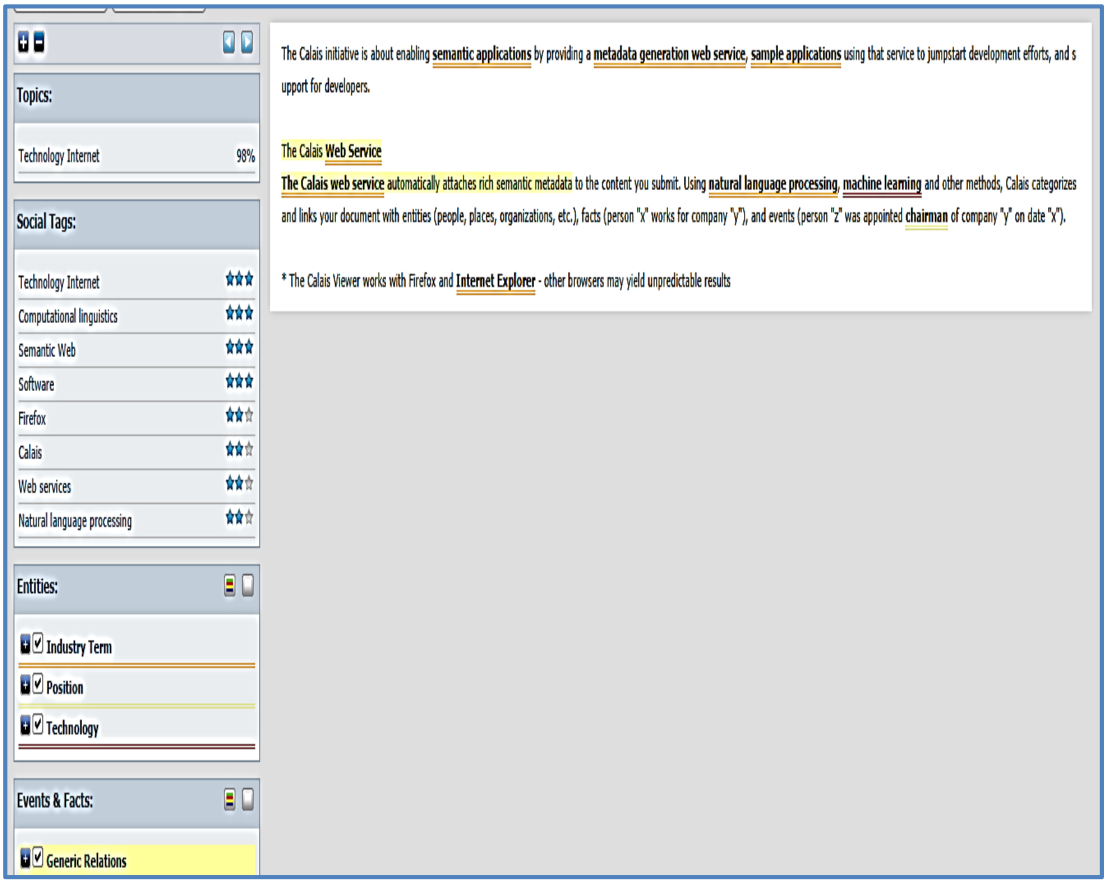


Figure 2-20: OpenCalais viewer interface

The main limitation for Zemanta and OpenCalais originates from their proprietary characteristic. There is a limitation on the number of requests for free use: For example, it is 1000 requests per day for Zemanta. In addition, users of these services cannot customise the systems to their needs or tailor the method of entity extraction to use different resources.

2.4.2.2 Wikipedia-based Approaches for Semantic Annotation

Most of the recent studies on semantic annotation and entity recognition in the last five years have utilised Wikipedia as a vast background knowledgebase that is available for use at no cost and is annotated by humans. The most commonly used knowledge bases in this case are DBpedia or YAGO (Yet Another Great Ontology) (Suchanek et al., 2008) because both are knowledge bases derived directly from Wikipedia and hence provide a direct mapping between an entity in the text and the related Wikipedia article through the use of URIs. These are ontology-based approaches that have roots in previous attempts to enrich text with tags and links extracted from Wikipedia pages (For example, (Milne and Witten, 2008; Hu et al., 2009; Banerjee et al., 2007)).

Entity recognition and linking approaches that depend on Wikipedia knowledge bases normally build a dictionary of tags (labels) for each URI using three sources: terms pages, disambiguation articles (for terms that have multiple meanings), and links within articles that links to other Wikipedia pages. The dictionary of labels is used for recognising an entity in the given text and identifies all potential URIs for that entity. If the entity is not found in the knowledge base, a special "no match" value is assigned to it. Then, the potential URIs are ordered, usually based on the context and a confidence value is assigned to each URI.

Entity mentions in the text could be disambiguated individually or jointly through the whole document. Normally, approaches that are based on Wikipedia utilise its pages coupled with similarity measures, such as cosine similarity (see section 2.1.3). This approach identifies the context of the entity mentioned in the text and match against the text in Wikipedia articles for each candidate URI (Mendes et al., 2011). Michelson and Macskassy (2010) have shown that this approach can be used to discover topics of interest on Twitter using a knowledgebase derived from Wikipedia categories to disambiguate and categorise entity mentions in tweets.

These Wikipedia-based approaches have been evaluated mainly using gold-standard news datasets and Wikipedia pages. However, since there is no gold-standard data sets for social media (especially online discussion forums), researchers have been testing and evaluating their approaches and algorithms using different case studies and the data sets gathered by the researchers themselves.

One of the most commonly used Wikipedia based knowledge bases is DBpedia and its semantic annotation system DBpedia Spotlight (Mendes et al., 2011). Spotlight is a free web service that can be used to semantically annotate text with URIs from DBpedia. Spotlight uses DBpedia ontology, which has thirty two top-level classes and 529 classes in total[10], as a target knowledge base. Spotlight is highly customisable since users can restrict which classes of the ontology are to be used for semantic annotation, either through the web interface or in SPARQL Protocol and RDF Query Language (SPARQL) queries. Spotlight first utilises a dictionary derived from Wikipedia in order to look up entity mentions in the text and return candidate URIs. Then, URIs are ranked using a TF/IDF model (see section 2.1): Each DBpedia instance (known as resource) is connected to a special document built from all the mentions of the entity in Wikipedia articles. These special documents are used to assign a confidence score to each URI in order to select the closest, most suitable annotation to the entity. Several studies , such as (Mendes et al., 2011; Rizzo et al., 2012) have shown that DBpedia Spotlight has out-performed commercial semantic annotation services, such as Zemanta and Open Calais (see 2.4.2.1) when tested on a gold-standard news data set.

[10] As of February 2014

2.5 Specifying Research Objectives

For the purpose of this research, and combining the suggestions from the domain ontology and semantic annotation field review, a goal for helping in the analysis of discussion forums could be shaped as follows:

"Can the combination of specific domain ontology and a general knowledgebase, in a semantic text mining method address the challenge of automatically analysing online discussion forums with minimum reliance on domain experts?"

Obviously, this goal contains many elements and phases of modelling and development. Therefore, it should be broken down in order to concentrate on one element (or step) at a time. Hereafter, the overall goal has been divided according to the method development phases:

1. This research is proposing an enhancement to the SEA approach to generate/derive a specific domain ontology in order to be used for annotating and analysing online discussion forums. The proposed approach should be able to produce the ontology using semantically related terms from a knowledgebase. This stage should not rely on domain experts, and an algorithm (automated software) should be able to obtain the required terms and their relationships. The ontological structure should be clarified from the generated concepts, taking into consideration the rich internal structure required and the weighted directional relations.
2. The proposed approach should then model the combination and utilisation of the derived ontology and a general knowledgebase, such as DBpedia in order to semantically annotate the online discussion forums. This combination is necessary for identifying hidden relationships in the text that were not

modelled by either knowledge bases. The extracted knowledge can then be used to improve and enhance the ontology for any future use.

3. The automatic annotation of the online forums enables the temporal visual analytics of the text in the forums. For example, a map of the important concepts, which were extracted from the forums, can be visualised in the form of temporal tag cloud. This helps to provide a high-level understanding of the discussed information in the forum to researchers and users as it makes it easier for them to find the topics that they are looking for.

This increased granularity of the research objectives suggests the steps of a viable approach. In the next two chapters, the modelling and evaluation of the suggested approach are conducted in order to investigate whether it will meet the research objectives.

Chapter 3: Generating a Domain Ontology

The conclusion from the analysis of the problem background and previous approaches in chapter two has shown that ontology based data analytic approaches where domain knowledge is important can be effective. The SEA ontology building approach had advantages for building domain ontology without access to experts, quickly and reliably. However, SEA was developed for producing ontology suitable for resource matching applications rather than text mining. In addition, the SEA's data source is no longer available (GoogleSets). This section explores a new source for the proposed approach that may be able to replace GoogleSets, but also be better focussed towards text analysis rather than resource matching. The following discussion explores enriching the generated ontology with links to a knowledge base that makes it suitable for annotating and analysing online discussion forums. This helps make the approach a better tool for non-domain expert users.

3.1 Process Configuration for Ontology Generation

This section describes in detail the techniques used to configure the building process of the diabetes ontology using a new knowledge source.

3.1.1 Knowledge Source Selection

The source of the ontology is the container of the domain knowledge. As discussed in section 2.3.1.2, the reuse of knowledge sources is highly preferred to produce ontology/terminology quickly and with less reliance on domain experts.

Traditionally, Google and other search engines have provided the users with domain-specific content either via their directories (such as Yahoo! directory) or via their adwords tools (such as the Google Adwords Keyword Planner). Search engine directories are very much like industry classifications, which often suffer from inadequate or unsatisfactory domain coverage as well as lack of meaningful relations. However search engines try to compensate for the limited relations in their directories by offering keyword generation tools, which provide users relevant concepts and terms through an analysis of the history of search queries by users of the search engine. Nevertheless, as these tools depend on the input of users, as discussed, they become very similar to the idea of building folksonomies, with the disadvantages that have been discussed in section 2.3.1: chaotic relationships accepted by the search engine may weaken the integrity of the domain knowledge.

Google used to provide GoogleSets for generating semantically related terms based on analysing its search index. SEA (Ma et al., 2014) utilised GoogleSets as a tool to produce relationships between starting keywords and the Google index as the knowledge source. It could be regarded as a word-clustering tool. However, Google discontinued GoogleSets in 2011 as part of shutting down parts of the Google Labs (Google, 2011) toolset. Although GoogleSets continued to be used in Google Docs for a while as part of the special “auto-fill” functionality, Google stopped it completely in 2014 version of Google Docs. In addition, GoogleSets did not provide links from the generated concepts to a knowledge base, which undermined its ability to be used for annotating the text data of online discussion forums, and its usability to perform word sense disambiguation (WSD) tasks. As a result, there is a need for a tool that provides semantically related terms to the starting keywords and links them to a knowledge source.

3.1.2 Semantic Relationship

The discussion of linguistic ontology in section 2.3.1.3 suggested that a semantic relation could be used to connect the starting keywords with the knowledge source.

This means that each starting word would be linked to associated concepts in the knowledge source based on semantic connections. These concepts are cluster(s) of terms in the domain, and are represented by the starting keywords. Generating these clustered concepts is usually carried out using two techniques, classification or categorisation:

- **Categorisation** is the procedure of dividing the world (text in our case) into sets of entities (Aas and Eikvil, 1999). The members of each set are *similar* in a specific context. Research in the categorisation area focuses on *concept formation* as a method of recognition.
- **Classification** is a process that includes assigning terms to mutually exclusive classes systematically, where each one of the terms belong to one class in a system, where the classes are not overlapping (Jacob, 2004).

In fact, categorisation is preferred to classification for this research, since it allows for overlapping concepts, which enhance the coverage over the domain. This facilitates the use of fuzzy concepts that connect the derived terminology to other terms within the domain and, more significantly, to the non-expert language and terms in the domain. Classification relationships usually result in gaps between the domain concepts as they emphasise the distinctiveness of the concepts, do not allow overlapping, and highlight restrictions and borders between concepts.

Categorisation includes a number of techniques that may be used for this research, but one method known as “Word Clustering” (EAGLES, 1998) uses directly the concept of co-occurrence in order to form semantic connections between concepts. Within this method, two sets of terms are processed into one category if they are “*similar*”. Concept similarity can be measured in two different ways:

- **Semantic Similarity**: Two terms are considered semantically similar if they are related by "virtue of their likeness" and can be substituted in a given context (Budanitsky and Hirst, 2001). For instance, in the sentence "I am reading a book", the term "book" can be replaced with "magazine" with no change in the sentence structure, and little change in the sense of the sentence. Hence, "book" and "magazine" can be considered as similar semantically.
- **Semantic Relatedness**: Two terms are considered semantically related when they co-appear often together in sentences and documents. Semantic relatedness takes into consideration any kind of relationships between the terms, including synonyms, antonyms and hypernyms[11] (Kulkarni and Caragea, 2009). For example, words "Eat" and "Lunch" are considered related terms as they usually occur together in the same context.

This research chose to utilise semantically related concepts, as terms that represent the same concept are more likely to be semantically related than semantically similar terms. Semantically similar terms are actually often covered by semantic relatedness: Replaceable terms could appear together in the same text, such as *breakfast* and *lunch,* while related words that appear together may not be mutually replaceable, such as *eat* and *lunch*. Semantic relatedness word clustering could offer the required binary measurement of the connections between concepts within the ontology.

OneLook (www.onelook.com) and its tools could be considered as a replacement to GoogleSets for generating related terms from initial seeding words. It is a search engine that indexes definition entries and dictionary websites across the internet as well as other sources such as Wikipedia articles. OneLook as a search engine allows

[11] Hypernym is a term used to represent a word that is a superordinate to other words. For example, ***Animal*** is a hypernym of ***Bird*** and ***Mammal***.

the user to customise the knowledge source that will be used to find the related results. It has indexed over a 1000 sources, including Wikipedia, Wiktionary and online medical dictionaries. The main search function will return a definition of the word searched for in all the indexed dictionaries. Other options include finding phrases with specific words in it. Unfortunately, none of these options is a useful provider of the required functionality for our research. However, OneLook contains what is called a "Reverse Dictionary", which allows the users to explore the related concepts to the search term, as well as generating a list of concepts in a given category. For example, searching the reverse dictionary for "*Green Fruit*" will return a list of terms that includes common known green fruit such as "*Kiwi*", "*avocadoes*", "*watermelon*", "*grape*" and "*apple*" and other less used terms such as "*Alligator Pears*" (another name for Avocado) or "*irvingia gabonensis*" (known also as *Wild Mango* or *African mango*).

The reverse dictionary is able to produce a list of words, phrases and abbreviations that are *related* to the input concept. However, it is not clear what type of relationship is produced by it. Therefore, an experiment was carried out to determine whether the output of OneLook is semantically similar or semantically related to the input terms. In this experiment, five general keywords from different domains were fed into the OneLook Tool and the output concepts were obtained. These testing keywords were chosen to represent general categories that are discussed by users on forums. The test seeding words were: *Sports*, *Insurance, Food, Medicine* and *Electricity*.

The results of all the five experiments showed that all of the OneLook tests (five out of five) generated a mixture of both semantically related concepts and semantically similar concepts, with the majority being semantically related. The results could be regarded as a semantic relatedness output since the two similarity types are not contradictory: semantically similar concepts are considered semantically related, which is why they might appear in the results. Hence, the “Reverse Dictionary” of OneLook was considered to be a semantic relatedness-based service that can connect the concepts to the indexed knowledgebase.

Another advantage of OneLook compared to GoogleSets is that the output of the reverse dictionary is linked with the parent knowledgebase and can be tracked back to its source(s). In addition, the reverse dictionary can provide related terms and concepts based on input categories, such as “Big Birds” and “Red Fruit” while GoogleSets did not provide such capability. This can help the ontology builder to choose starting words that are more representative of the domain without significant help from domain experts.

Another difference between OneLook and GoogleSets lies in the fact that GoogleSets takes as an input multiple keywords (up to five keywords) in separate text boxes, while the OneLook input box is similar to a search engine’s text box. Therefore, in order to test the impact of the number of input keywords on the output, a software program to perform an “intersection” between the result sets of the starting keywords was developed. This program takes as an input a maximum of five keywords (Similar to GoogleSet) and calls the OneLook Reverse Dictionary service to obtain the semantically related results. The program was used to test the different configurations of OneLook Reverse Dictionary in order to identify the optimal configurations suitable for the research goals.

OneLook (Figure 3-1) has several options to control the output of the results. For example, it can be customised to use just limited number of indexed sources. The effects of using the options and the customisation on the generated terms were not very clear. This required researching the different options so they can be set to the optimal configuration to generate the best results.

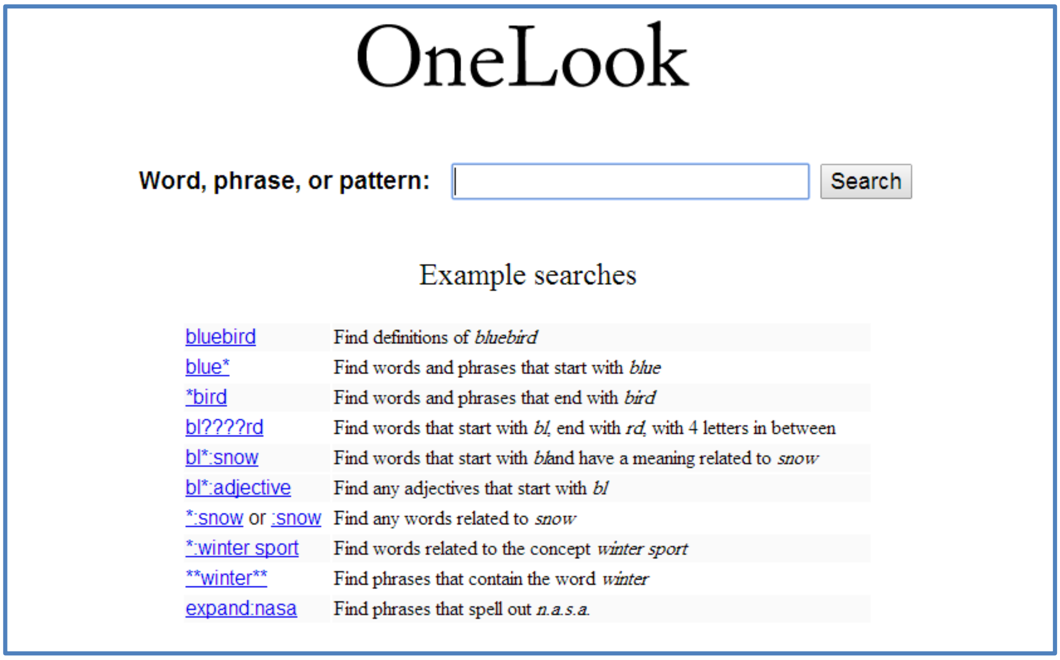


Figure 3-1: The OneLook Interface

3.1.3 Ontology Generation Experiment Configuration

OneLook has a number of options that could tune the results. In this section, the exploration of options to find the optimal configuration to generate the most valid results is described. Apart from enabling the choice of the knowledge source, OneLook is different from GoogleSets in that it does not stipulate a limited number of input words. Normally, the reverse dictionary is configured to find related concepts for either one concept, such as “Diabetes”, or for a category, such as “Black Birds”. The result from the Reverse Dictionary is a large domain space that needs further processing. In addition, the required or optimal number of starting keywords needs to be defined. In order to identify the optimal configuration, the following points need to be investigated:

1. **The number of input keywords**: Would a change in the number of starting keywords make a difference in the output? What is the *optimum* number?
2. **Keywords' Impact on Domain Focus**: How will the chosen keywords affect the output predictions with respect to the target domain? Can this number be used to make the results more domain-directed?
3. **Starting Keywords' Order**: Does the output change based on the order of the input keywords? If yes, what is the best order of these keywords?

3.1.3.1 Configuring the number of starting keywords

The number of input keywords may have two main impacts over the output predictions: on number of generated concepts (Quantity) and on the subject area focus (Quality).

3.1.3.1.1 "Quantity of Output" Effect

Five key words were chosen as the input for this experiment and they were run through the OneLook five times (single seeding keyword, two seeding keywords, three seeding keywords, four seeding keywords and five seeding keywords) in order to check the output of the OneLook platform (Figure 3-2). The output shown in figure 3.3 revealed three main findings:

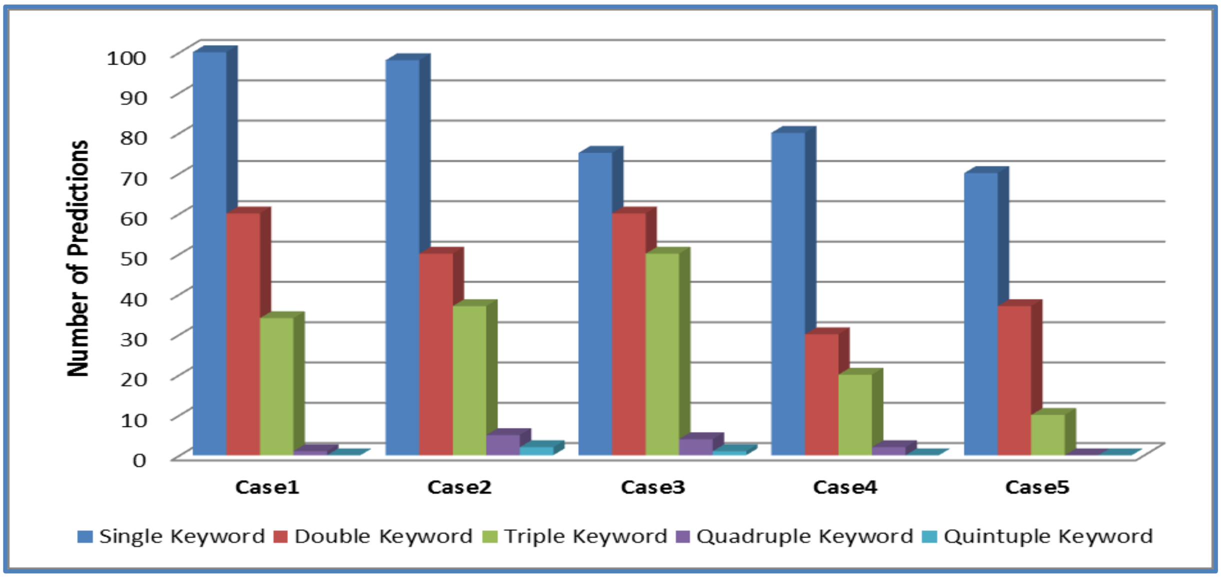


Figure 3-2: Relationship between number of seeding keywords and output concepts

1. The results suggested that the number of output words was inversely proportional to the number of seeding keywords. The bigger the number of input keywords, the more constraints the intersection had on the semantically related output. Consequently, there were fewer results because of the increased constraints created a more focused domain.
2. A big fall in the number of outputs was experienced when more than two terms were used as seeding keywords, especially in case 1. It seems that each new word restricts the search space that the OneLook reverse dictionary utilised. The decreased number of output terms suggests that the more seeding keywords OneLook had, the smaller the number of results would be generated.
3. Single and paired seeding keywords are the ones that generated more concepts in all the cases compared to the other three configurations.

It seems that either single or paired keywords could be the optimal setting required to produce the required semantically related concepts. However, a large number of output terms does not automatically mean the expected domain focus has been achieved. Hence, additional experiments were conducted in order to test the effect on the quality of output from the domain focus' point of view.

3.1.3.1.2 "Quality of Output" Effect

Single and paired seeding keywords were the cases used to study the impact on the "quality" of the output concepts. In this test, the word "quality" reflects the focus level of the studied domain and the capability of the approach to produce rich statistical data regarding the importance of various output terms.

When using keywords as seeds to build an ontology, it is predictable that more domain-focused terms are going to be generated if double (paired) keywords were employed to limit their respective domains and produce a concentrated mutual domain. Most of the single input keywords would generate terms in various subject areas (facets) with less restrictions. Hence, double/paired keywords are more likely to outperform single keywords in generating focused concepts (Figure 3-3). However, this hypothesis needs to be validated since OneLook.com's Reverse Dictionary has not been tested from domain focus perspective.

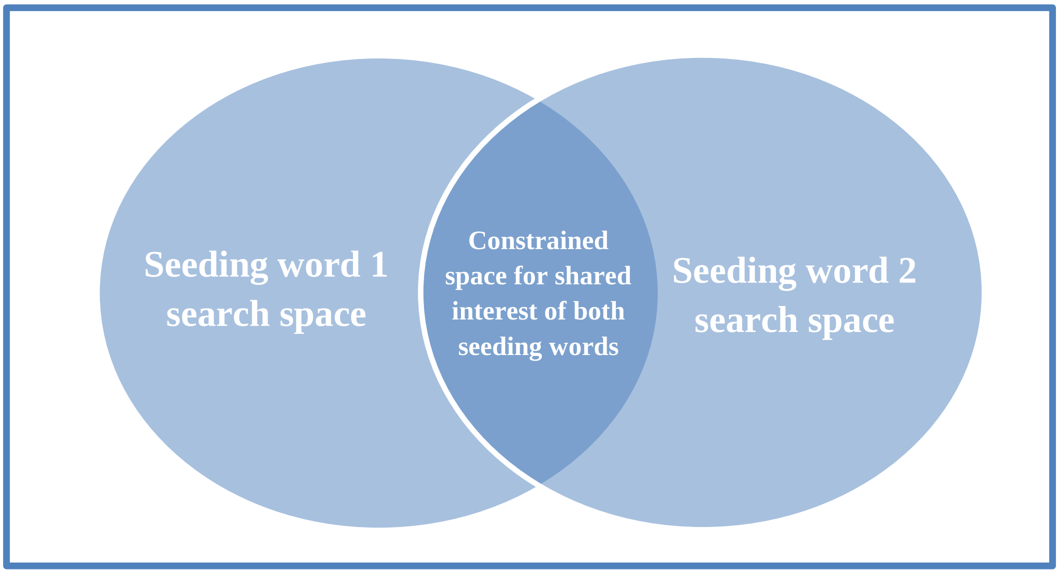


Figure 3-3: Double seeding keywords constrained search space

Therefore, the purpose of the *quality* tests is to evaluate the output of the Reverse Dictionary when double keywords are used, and compare it to the output of using a single keyword. Hence, a generalised decision could be made whether to use single or paired keywords as the starting point for generating an ontology. For this experiment, three medical terms were chosen to be the input for our experiment, ranging from general concepts to specific ones:

- **Cancer Treatment**: A high level category in the medical area that is publicly discussed and likely to appear in everyday conversations and news;
- **Plantar Fasciitis**: A medical concept that describes a very common disorder and may appear in general conversations;
- **Sacral agenesis**: A congenital disorder that is only used in a professional medical context.

OneLook's Reverse Dictionary was used to produce the related concepts. Unfortunately, the number of output concepts and relationships did not reflect any specific domain or produce results that are significant statistically. In order to overcome this problem and produce more domain-specific concepts, a "Snowball Sampling" approach (Handcock and Gile, 2011) was adopted. In this approach, the output of the first experiment are exploited as the seeds for the second one and so on.

Snowball sampling is known in social science and statistics to generate meaningful distribution as a network, and consequently social network analysis (SNA) methods can be used to discover more about this network and its members (concepts in our case). In fact, the use of snowball sampling to generate the domain-specific concepts not only produced a network, but also generates "occurrence degree"[12] for the nodes

[12] The frequency of occurrence

of the network (concepts). Some concepts occurred (appeared in the output) repeatedly and hence had a higher degree of occurrence, while other concepts occurred only once in the output.

The three terms (Cancer Treatment, Plantar fasciitis, and Sacral agenesis) were used as seeding keywords for the "paired keywords" experiment. The first level experiment was identical to the previous single experiment regarding, the second round experiment was different as the output of the first round were coupled with the input from the first round. For example, if *Diabetes* was the input for the first round and the Reverse Dictionary produced *Insulin* and *Pancreas* as the output, then the paired input for the second round would be:

- "*Diabetes*" and "*Insulin*"
- "*Diabetes*" and "*Pancreas*"

The results of the experiments with both single and double (paired) keywords are shown in Table 3-1 and Table 3-2.

Table 3-1: Single keywords experiment

Seeding Keywords	Unique Concepts	Total Concepts	CFM Ratio	Long Tail (Total)	Long Tail (Percentage)
Cancer Treatment	3023	4144	1.37	2408	79.66%
Plantar Fasciitis	447	905	2.02	135	30.20%
Sacral Agenesis	516	1051	2.04	283	54.84%

Table 3-2: Paired keywords experiment

Seeding Keywords	Unique Concepts	Concepts (Total)	CFM Ratio	Long Tail (Total)	Long Tail (Percentage)
Cancer Treatment	101	503	4.98	12	11.88%
Plantar Fasciitis	13	49	3.77	1	7.69%
Sacral Agenesis	17	106	6.24	1	5.88%

The data in the tables shows the following:

- "**Unique Concepts**": This number represents the output terms without their frequency. It treats the repetition as the same concept.
- "**Total Number of Concepts**": This number reflects the number of all the generated terms regardless of their repetition. It represents the sum of the terms' occurrence frequency and at the same time is a measurement of the relations between the terms produced by this experiment, as there is a semantic connection between each output term with the seeding word that generated it.
- "**Concepts Frequency Mean (CFM)**": This number is calculated as the ratio of the previous two numbers:

CFM = Total Number of Concepts/ Number of Unique Concepts

This number represents the *"The Average Nominations"* any concept (node) has from the other concept (nodes) of the same *network*. The more connections a particular node (concept) has, the more likely that it represent the network (Gjoka et al., 2011). Consequently, the CFM measures the degree to which the output represents a focused domain.

- **"Long Tail (Total)"**: This is the number of all the concepts that were generated only once in the output of the experiment. Since it was only produced by one input concept in the network, this type of output is not considered a representative concept within the target output domain (an *outlier*). Therefore, the *longer* the tail is, the less focused the output domain is.
- **"Long Tail (Percentage)"**: This percentage can therefore represent the degree of how *scattered* the output domain is. The bigger the percentage is, the less focused the resultant domain is. Nevertheless, the concepts in this long tail may be seen as the "links" between various domains, and consequently are important when building an ontology that serves multiple domains.

When comparing the output of both single and paired keywords experiments, the following distinctions were recognised:

a) The output of paired keywords experiment was a more focused network (domain) from both CFM and long tail percentage's point of view:

   - In the "Cancer Treatment" experiment, pairing up the starting keywords resulted in a CFM value that is more than a triple of what the single keyword experiment produced (4.98 compared to 1.37). In the "Sacral Agenesis" domain, the CFM tripled to 6.24 from 2.04, while the domain focus doubled in "Plantar Fasciitis" experiment from 2.02 to 3.77. It can be concluded that the paired keywords experiment produced superior results and managed to increase the level of the domain focus in all instances irrespective of the expected level of domain focus that the starting keywords carry naturally.

- As to the long tail percentage, pairing keywords had an advantage when compared with single keywords configuration, as it reduced the number of the long tail predictions in all three experiments as Table 3-1 and Table 3-2 show. For example, “Sacral Agenesis” output’s long tail had only a value of 5.88% compared with 54.84% in the single keyword experiment. The small percentage of long tail terms shows that using paired keywords as an input produced a more focused domain compared with the single keyword configuration.

b) The “Cancer Treatment” experiment shows that the paired keywords configuration was particularly helpful to generate a better concentration of the resultant predictions in what was expected to be a less-focused domain. In this experiment, the resulting network was more likely to produce the least accurate output, because of the general nature of the term. In fact, the single keyword configuration showed that the “Cancer Treatment” output domain is the least-focused amongst the three experimental domains – with a CFM value of 1.37 and a long tail percentage of 79.66%. The paired keywords experiment pushed the CFM up to 4.98 with only 11.88% long tail percentage. A closer look at the output concepts showed that the use of single keywords produced many concepts that are not specifically “treatments” but they were semantically related terms in the medical area. For example, terms such as “*surgery*”, “*response*”, “*Paget's disease*” and “*pioneer*” were produced in the single keyword experiment. These words brought back their related terms in their domains in the second round experiment as single words and, consequently, generated non-cancer-treatment related concepts. However, when these terms were paired with the "cancer treatment" concept, their output was constrained, so the output was related concepts to the "Cancer Treatment" domain.

c) The pairing of keywords has also enhanced the concentration of the domains that are naturally more focused. In "Plantar Fasciitis" experiment, the focus level has been improved from around 2.02 to just 3.77. This is probably a result of the characteristics of Plantar Fasciitis's domain, which is naturally focused. An unusual case of this scenario is another keyword (*Sarcoma*) output that may offer an interesting example for this situation. In the first round experiment, "Sarcoma" has generated fifteen terms: ***Sarcomata, Sarcomatous, Osteosarcoma, Chondrosarcoma, Kaposi's sarcoma, Neurosarcoma, Leiomyosarcoma, Liposarcoma, Myosarcoma, Osteogenic sarcoma, Sarcoid, Sarcosis, Carcinosarcoma, and Sarcomas*.** All the fifteen terms were used in the second round experiment and the result of the experiment (except for "***malignant neuroma***") was the same fifteen terms (with varying frequency) in both single and pairing experiments. This implies that the fourteen words belong naturally to a well-defined and restricted domain (Figure 3-4). Such a level of domain focus will yield the same result regardless of how many rounds are conducted, or whatever the keyword number configuration is. It can be concluded that genuinely highly concentrated domains may not benefit from pairing the starting keywords as much as the less concentrated ones.

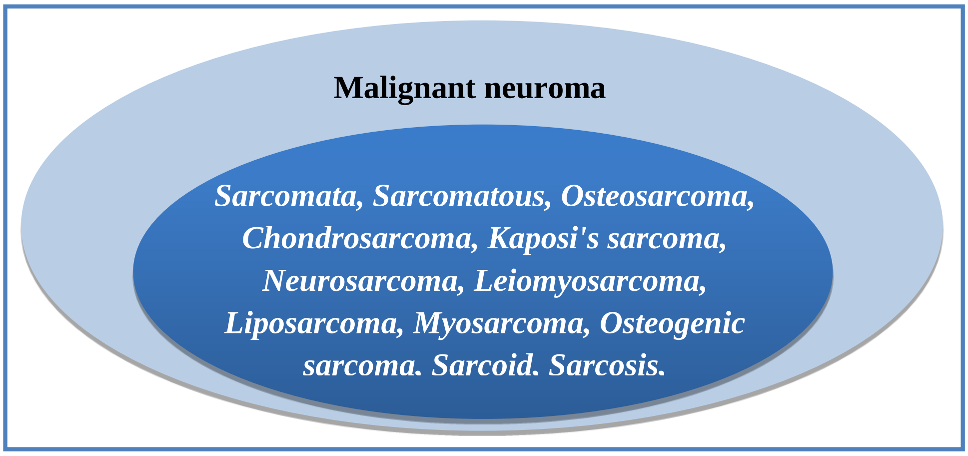


Figure 3-4: Sarcoma output domain focus

From a theoretical perspective, going through one round of generating semantically related concepts may only produce a set of terms that does not have any differences regarding their occurrence, network position, or connections between the terms. However, with the two round *snowball sampling* technique, the terms start to have a statistical significance (frequency) as well as various asymmetric relationships within the network that underlines a terms' importance in the domain. This clarifies the differences between various terms. For some domains, a third round experiment might be needed to "amplify" the differences within the domain between the various terms (Figure 3-5). These clearer distinctions between terms were valuable when conducting the analysis of the ontology network compared with the more-even statistical distribution of the single starting keywords output.

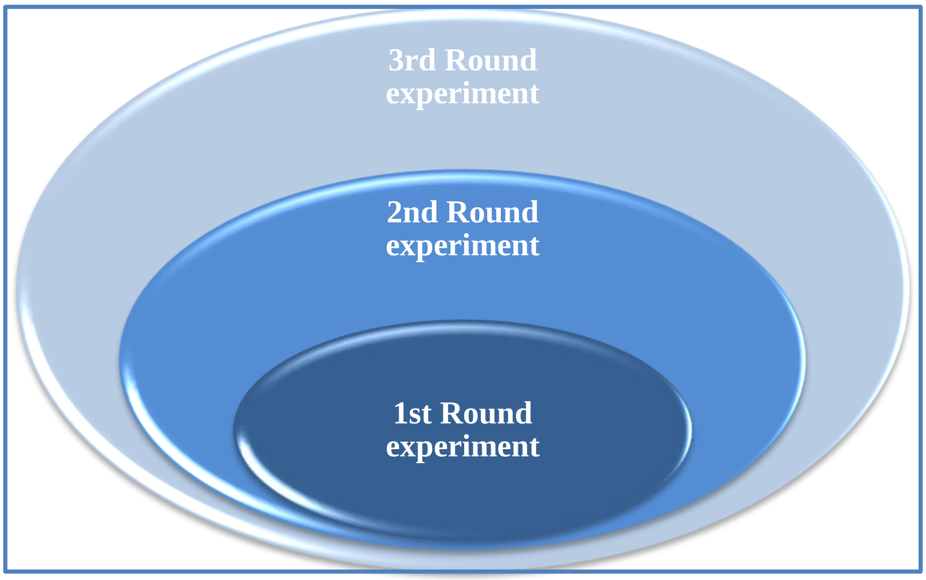


Figure 3-5: Multiple round experiments

This indicates that pairing the starting keywords would generate better results when compared to single keywords from the statistical significance point of view. In fact, the output of the experiments indicates that pairing starting keywords improves both the quality of domain focus as well as the quality of the generated terms for additional examination. Hence, paired keywords were chosen as the seeding configuration for additional analysis.

If paired keywords were selected as the configuration of the starting number of keywords, a question that needs to be answered is: Will the order of the keywords affect the output of the ontology?

3.1.3.1.3 Impact of Keywords' Order

Since the paired keywords' output consists of the intersection of the two sets of results of single words, the order of starting keywords has no effect whatsoever on the predictions generated. Since no difference in the output appeared from all testing experiments, this indicates that OneLook.com paired keywords setting is "*order insensitive*".

The experiments conducted for the SEA methodology (Ma et al., 2014) suggested that one pair of starting keywords is not enough to generate a desirable domain specific output and concluded that three pairs of starting keywords should be used. This indicates that further experiments in this research are required in order to discover the ideal number of starting paired keywords when using the Reverse Dictionary.

3.1.3.2 Domain Sensitivity Configuration (Starting number of keywords)

The SEA methodology (discussed in section 2.3.1) argued the need for three pairs of starting keywords in order to avoid the input misleading the output direction when using GoogleSets. As to this research, the discussion in 3.1.3.1 showed that the output of the experiments using the Reverse Dictionary of OneLook.com was susceptible to the level of domain focus presented by the starting keywords. In fact, the chosen starting keywords must be highly illustrative and representative of the domain, as well as being more related to other meanings within the target domain terms.

The utilisation of domain experts may be needed to choose terms that could produce a better level of domain focus. However, traditional methodologies for generating ontology had the disadvantage of depending heavily on the domain experts, and this is a major issue that this research is trying to overcome. Even if we obtain an involvement from domain experts, the output terms may still be inadequate and susceptible to the bias of experts. They could have weaknesses in their knowledge domain coverage (which is largely why normal approaches usually require multiple experts) and/or enforce their opinions on topics in the domain. Therefore, it is important to have a configuration setting to tackle the sensitivity of the domain regarding the starting number of paired keywords. This helps to cut the dependency on the domain experts and at the same time offers a "fault tolerance" mechanism capability for the input starting keywords.

Conceptually, the increase in the number of starting paired keywords can lead to improvement in the tolerance ability of the proposed process. Hence, there is a need to define the smallest possible number of starting seed pairs of keywords to achieve this. The main reason behind this decision is the vulnerability of the methodology to the possible misdirection of the domain when using only one pair of keywords as the input for the first round experiment.

1. **One pair of keywords**: Since there is only one pair in this configuration, the fault tolerance ability is limited. Any ambiguity in the selection could lead to misdirection in the concept search. Figure 3-6 shows that combining the keywords K & H could lead to two separate domain networks (M-centred domain and P-centred domain) that do not have any common concepts other than K and H.

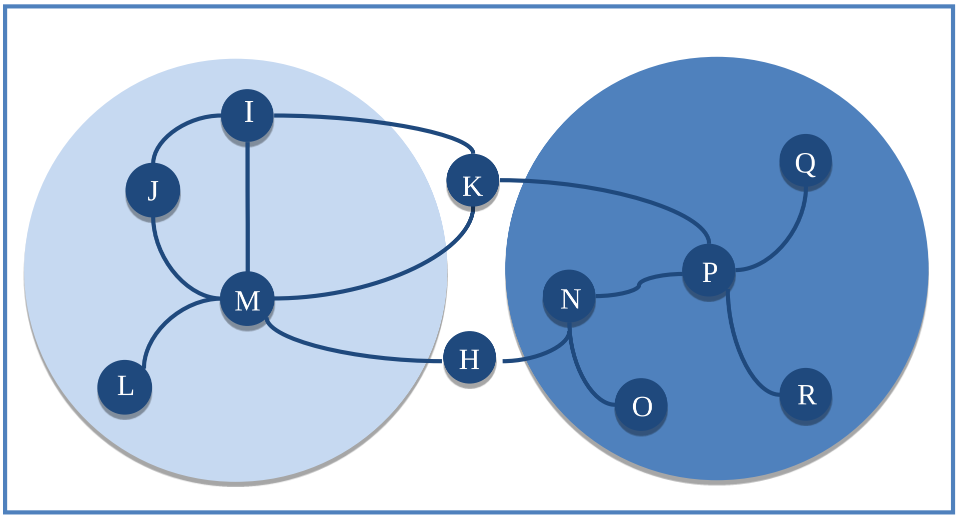


Figure 3-6: Misleading pair of keywords

2. **Two Pair of Keywords**: The use of a second pair of keywords would be preferable when one pair does not focus on the target domain, though still producing domain-specific concepts. However, the use of two pairs may lead to concepts around two separated domains. Figure 3-7 shows an extreme example where one pair of keywords has not generated any domain-specific concepts at all.

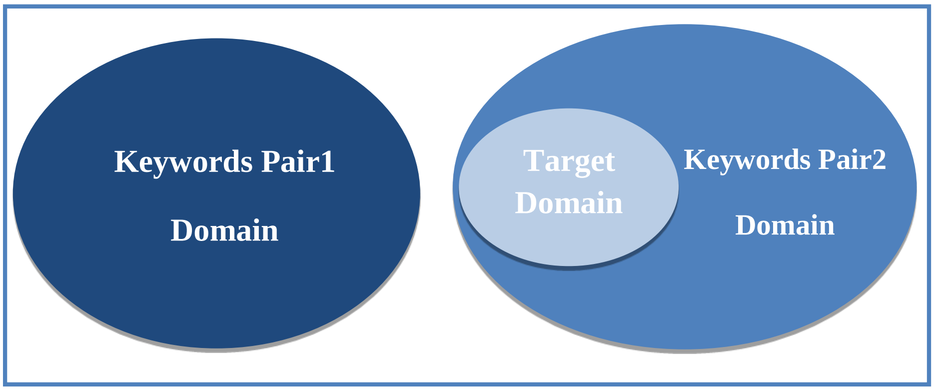


Figure 3-7: Complete separation of two domains generated by two pairs of keywords

The output of the experiment resulted in two completely separated terms divisions, with no shared concepts. In such an extreme case, the output ontology is not domain-specific and requires more involvement from domain experts in order to guide the pruning of the output to target the required domain.

3. **Three Pairs of Keyword**: In this case, if one pair of starting keywords were misleading and resulted in a separate unrelated domain different from the required one, the other two pairs will produce concepts that still belong to the target domain. The overlap between the remaining predictions can create a high frequency of emergence of these output concepts compared with the output of the misleading pair of keywords (Figure 3-8). The overlapping predictions therefore will have a lead over the misleading ones statistically.

For this type of situation, the generation of ontology concepts could yet be considered worthwhile and, more importantly, there is no need for the involvement of the domain experts compared with the previous setting.

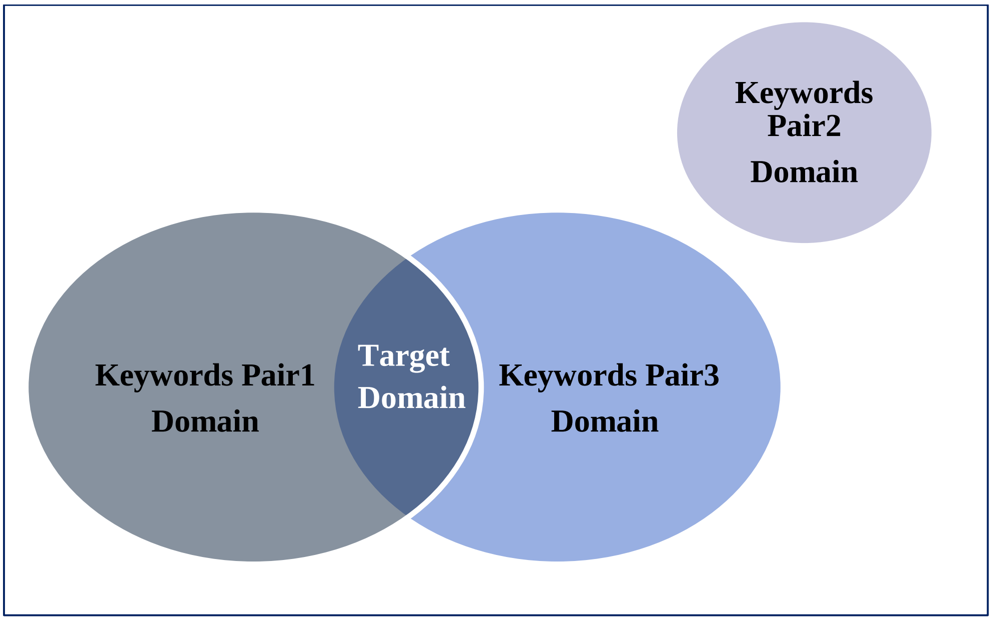


Figure 3-8: Three pairs of keywords scenario with a misleading pair

The "starting number of keywords" experiment has shown that there could still be a dependency on domain experts. However, this is restricted to just suggesting two or three keyword seeding pairs. The experiments show that even if the domain expert suggests a misleading pair of keywords at the beginning of the ontology corpus building process, the proposed approach is able to tolerate this flaw and still produce a domain ontology concept mapping. Thus, we can conclude that ideally it is best to start with three domain keyword pairs in the seeding words for a OneLook based methodology.

3.1.3.3 Summary of Ontology Building Configuration

To summarise the output of the previous experiments, domain ontology building using the Reverse Dictionary of OneLook should be carried out as follows:

1. Obtain the starting keywords as three pairs of terms either from the domain experts or by a preliminary experimentation based on the target application;
2. Two rounds of concept generation experiments should be conducted, a third round is optional depending on the output predictions from the first two;
3. For the first round of experiments, the order of feeding the pairs of keywords is irrelevant. This is also true for the following rounds.

These steps emulate the first two stages in the approach proposed by SEA. In addition, they execute a part of the third process: Identifying the semantic relations and paths between the starting pairs of keywords and the knowledge base of OneLook through extracting the semantically related concepts from the reverse dictionary.

These configurations have sought to optimise the settings of the OneLook Reverse Dictionary to be used for producing the ontology corpus. Nevertheless, carrying out the building process for the ontology corpus manually is not practical for two reasons:

1. The number of concepts that the rounds generate will probably be too large to handle manually.
2. Keeping track of the semantic relationships between the starting keywords and the output predictions manually may result in errors. Therefore, a set of algorithms have been defined and a software program has been developed in order to automate the ontology building process. The ontology building process is discussed in section 3.2.

These steps do not identify the “root” concepts, since the Reverse Dictionary does not provide a structure for any given domain. Therefore, the ontology corpus has to be constructed in order to identify the “root” concepts and then an ontological analysis can be conducted in order to clarify the structure of the ontology.

3.2 Ontology Building Process

The Snowball Sampling technique (Handcock and Gile, 2011) has enabled building a corpus of the ontology and capturing an adequate number of terms in the target domain in order to create the ontology. The required process of building the ontology using three pairs of starting keywords is incorporated into an algorithm in order to facilitate automating it via a computer program.

3.2.1 First Round Corpus Building Experiment

The building process accepts three pairs of keywords as input and brings back the output generated from the reverse dictionary. The following formula represents the function of obtaining the output of a paired keywords x and y:

$$\boxed{R_{(x,y)} = f_{RD}(x,y) = \left\{w_1^{x,y}, w_2^{x,y}, \dots, w_{n_{x,y-1}}^{x,y}, w_{n_{x,y}}^{x,y}\right\}}$$

Equation 3-1

Function $f_{RD}(x,y)$ is the number of steps to capture the Reverse Dictionary's output using x and y as its input (seeding words). Result $R_{(x,y)}$ is the set of terms from $w_1^{x,y}$ to $w_{n_{x,y}}^{x,y}$ that was generated from both keywords (The superscript "*x, y*" is an indicator of the output set, while $n_{x,y}$ represents the number of generated terms by function $f_{RD}(x,y)$).

In the proposed methodology, S1&S2, S3&S4, S5&S6 are the three pairs of starting keywords that are used in the first round experiment, which are chosen to represent the domain/subject area. These pairs can be supplied either by domain experts or from an old existing ontology. This round can be formalised as follows:

$$R_{(s1,s2)} = f_{RD}(s1, s2) = \{w_1^{s1,s2}, w_2^{s1,s2}, \dots, w_{n_{s1,s2-1}}^{s1,s2}, w_{n_{s1,s2}}^{s1,s2}\}$$

$$R_{(s3,s4)} = f_{RD}(s3, s4) = \{w_1^{s3,s4}, w_2^{s3,s4}, \dots, w_{n_{s3,s4-1}}^{s3,s4}, w_{n_{s3,s4}}^{s3,s4}\}$$

$$R_{(s5,s6)} = f_{RD}(s5, s6) = \{w_1^{s5,s6}, w_2^{s5,s6}, \dots, w_{n_{s5,s6-1}}^{s5,s6}, w_{n_{s5,s6}}^{s5,s6}\}$$

Equation 3-2

For example, $w_6^{s3,s4}$ is the sixth output term produced by $f_{RD}(s3, s4)$ using the paired keywords S3&S4 (Detailed algorithm in Appendix A).

3.2.2 Second Round Snowballing experiment

The output of the first round experiment was collected and paired with the input starting keywords to form the new input for the second round. The resulting pairs were entered to the reverse dictionary in order to obtain more semantically related terms. For instance, if the first round input keyword S5 was paired with its ninth output prediction from the first round $w_9^{s5,s6}$ in $R_{(s5,s6)}$, then the pair would be one of the starting keywords to the second round snowballing experiment to produce a new set of terms $R(s5, w_9^{s5,s6})$ that has $n_{5,\,9,\,s5,\,s6}$ terms from $w_1^{5,9,s5,s6}$ to $w_{n_{5,9,s5,s6}}^{5,9,s5,s6}$:

$$R(s5,\, w_9^{s5,s6}) = f_{RD}(s5, w_9^{s5,s6}) = \left\{w_1^{5,\,9,\,s5,s6},\; w_2^{5,\,9,\,s5,s6},\; \dots,\; w_{(n_{5,\,9,\,s5,s6}-1)}^{5,\,9,\,s5,s6},\; w_{n_{5,\,9,\,s5,s6}}^{5,\,9,\,s5,s6}\right\}$$

Equation 3-3

The same formula was applied to all the combinations of any of the first round output terms with either of the keywords that produced them. Therefore, if the first round produced $n_{x,y}$ unique terms from the pair (x, y), then the second round experiment will have 2* $n_{x,y}$ input paired keywords ($n_{x,y}$ terms combined with each of x and y). Therefore, for S1 and S2, the second round experiment would be formalised as follows:

$$R\left(s1,\ w_1^{s1,s2}\right) = f_{RD}\left(s1, w_1^{s1,s2}\right) = \left\{w_1^{1,\,1,\ s1,s2},\ w_2^{1,\,1,\ s1,s2},\ \dots,\ w_{\left(n_{1,\,1,\ s1,s2}-1\right)}^{1,\,1,\ s1,s2}, w_{n_{1,\,1,\ s1,s2}}^{1,\,1,\ s1,s2}\right\}$$

$$\vdots$$

$$n_{1,2}$$

$$\vdots$$

$$R\left(s1,\ w_{n_{1,2}}^{s1,s2}\right) = f_{RD}\left(s1, w_{n_{1,2}}^{s1,s2}\right) = \left\{w_1^{1,\,(n_{1,2}),\ s1,s2},\ w_2^{1,\,(n_{1,2}),\ s1,s2},\ \dots,\ w_{\left(n_{1,\,(n_{1,2}),\ s1,s2}-1\right)}^{1,\,(n_{1,2}),\ s1,s2}, w_{n_{1,\,(n_{1,2}),\ s1,s2}}^{1,\,(n_{1,2}),\ s1,s2}\right\}$$

and

$$R\left(s2,\ w_1^{s1,s2}\right) = f_{RD}\left(s2, w_1^{s1,s2}\right) = \left\{w_1^{2,\,1,\ s1,s2},\ w_2^{2,\,1,\ s1,s2},\ \dots,\ w_{\left(n_{2,\,1,\ s1,s2}-1\right)}^{2,\,1,\ s1,s2}, w_{n_{2,\,1,\ s1,s2}}^{2,\,1,\ s1,s2}\right\}$$

$$\vdots$$

$$n_{1,2}$$

$$\vdots$$

$$R\left(s2,\ w_{n_{1,2}}^{s1,s2}\right) = f_{RD}\left(s2, w_{n_{1,2}}^{s1,s2}\right) = \left\{w_1^{2,\,(n_{1,2}),\ s1,s2},\ w_2^{2,\,(n_{1,2}),\ s1,s2},\ \dots,\ w_{\left(n_{2,\,(n_{1,2}),\ s1,s2}-1\right)}^{2,\,(n_{1,2}),\ s1,s2}, w_{n_{2,\,(n_{1,2}),\ s1,s2}}^{2,\,(n_{1,2}),\ s1,s2}\right\}$$

Equation 3-4

The same process would be applied to the entire output from the first round experiment, and the new keywords will be combined to generate more domain terms from the reverse dictionary.

As discussed in 3.1.3.1, if the second round experiment did not produce enough terms and/or there was no clear “statistical trend”, a third round experiment would be carried out based on the output of the second round.

The ontology building process, stores the input keywords, the results and the relationships amongst the terms. The resulting concepts and their semantic connections have created a “domain network”, which has similar characteristics to various social networks. Social Network Analysis (SNA) techniques can be used in conducting the "*ontological analysis*" for the resulting network. These techniques are used in this research to analyse the ontological structure of the output concepts.

3.3 Network analysis of the ontological structure

This analysis represents the final phases of the ontology building process and it contains three main sub-processes:

1. Defining the Root, i.e. the network representatives;
2. Structure Learning: Quantifying and refining connections between the *roots* and the rest of the domain concepts;
3. Ontology Pruning: Concepts Clustering and defining the *borders* of the network.

The analysis of the network will enhance the applicability of the ontology in computer systems and the ability of shaping the experiment output from multiple points of view to fit various applications in the domain. The network analysis starts with the traditional centrality calculation for each output term in the network in order to derive their "*social network position*".

3.3.1 Degree Centrality

By building the ontology network using the snowball sampling, the nodes of the network that appeared in the results (nominated) more frequently than other nodes can be considered as more "*central*" and more representative members of the domain.

For analysing the ontology network, the centrality measure in this research refers to "Degree Centrality" as an analysis tool from Social Network Analysis (SNA). It is regarded as one of the most significant techniques in SNA to discover the "*primary*" nodes within the network.

Nodes that have more ties to other nodes in the network have an advantageous position, and considered more central in the network. To put it simply, the degree of the network member is equal to the number of connections the member possesses. In a directed network graph, it is important to distinguish between in-degree and out-degree centrality (Hanneman and Riddle, 2005a). If a node has more in-bound ties, it is considered *prominent*, while nodes with high number of out-bound ties are considered *influential*. For the ontology network, focus is more on the in-degree centrality rather than out-degree centrality, since our concern is on the important *actors* in the network.

3.3.1.1 In-Degree Centrality Calculation

The output of the experiments described in the "corpus ontology building process" (section 3.2) was 2*n collections* of terms, in the second round – where **n** is the unique number of terms from the first round.

In order to calculate the centrality of a particular term (*t*) in this network structure, the algorithm has to go through all sets of output from the last round and get all the possible connections between the target term (*t*) and the starting keywords. Therefore, the calculation would have the following two steps:

1) Checking whether *t* exists in every set of output terms (C), taking into consideration that *t* was not a seeding keyword for C. This is captured via the function $f_E(t, C)$ which is defined as follows:

$$f_E(t,C) = \left\{\begin{matrix} 1, t \in C \\ 0, t \notin C \end{matrix} \middle| f_{RD}(t,k) \neq C\right\}$$

$$Where\ C = \forall C_{(w_{pi},w_{pj})} \mid 1 \leq i \leq j \leq n$$

$$And{:}\ t \in \{w_{p1}, w_{p2}, \cdots, w_{pn}\}, k \in \{w_{p1}, w_{p2}, \cdots, w_{pn}\}\ and\ t \neq k$$

Equation 3-5

2) In the second step, the centrality for the term t will be defined as the aggregation of the function $f_E(t,C)$ for all the collections C:

$$f_{Cen}(t) = \sum_{i,j} f_E(t, C_{(w_{pi},w_{pj})}) \quad | \quad 1 \leq i \leq j \leq n$$

Equation 3-6

The centrality function $f_{Cen}(t)$ generated various degrees of centrality as a distribution of the changes (Figure 3-9).

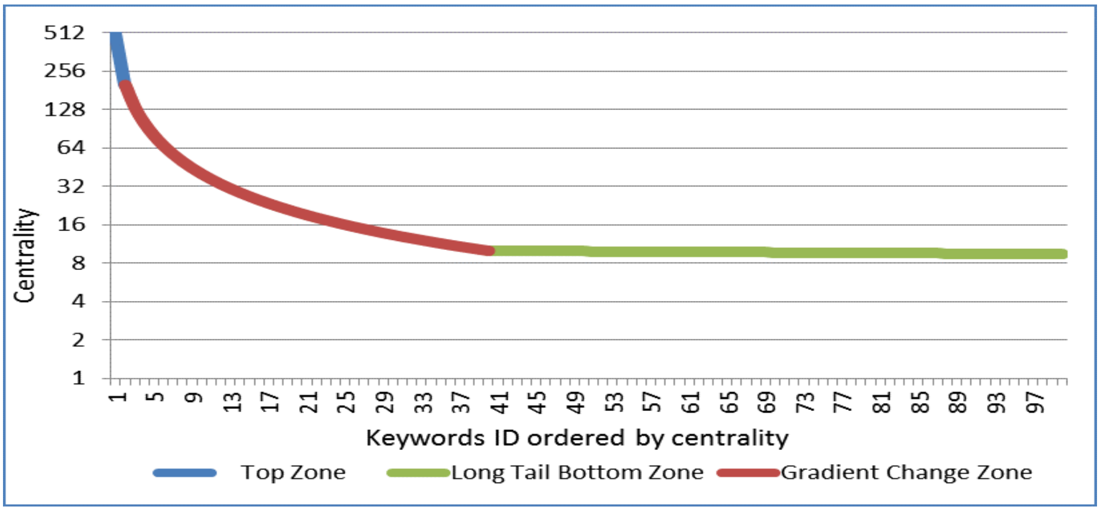


Figure 3-9: Tri-sectional Centrality Distribution Trend

In this kind of distribution, three notable sections can be differentiated:

1. **Top Section**: This fully connected section contains terms with high degrees of centrality. These terms were the most popular in the network, as they were the highest *nominated* concepts by other network members. These terms were the

concepts that define the subject area. Statistically, these concepts in the top section have been generated more often than any other concept in the network.

2. **Gradient Change Section**: This section contains concepts that are still highly connected, yet their degree centrality value was significantly lower than the top zone. Most of the concepts in this section were connected to top "rapid" section concepts. These concepts are not as important as the top section ones from a domain representation perspective, but they are complementary and closely connected.
3. **Long Tail Section**: This section consists of concepts with low degree of centrality with most concepts having a value of one (This value means that they were generated once in all the output collections). The concepts at this level do not essentially commonly associate with the concepts in the top or gradient change sections, but are connected to them in a specific context. They could be beneficial to this research as these concepts may play the role of connectors to other knowledge domains.

Degree centrality could find concepts that are prominent in the domain (top zone concepts) within the ontology network, but it fails to take into consideration the direct immediate connections that a node has to other members of the network. For instance, one node might have connections with a significant number of other nodes in the network, yet these nodes could be disconnected from the rest of the network: i.e. the node is only central within a local zone in the network (Hanneman and Riddle, 2005a). Therefore, more analysis is required in order to discover the relationships characteristics of various concepts within the network. Hence, relationship examination becomes an important issue in the next phase of analysis. Two main types of analysis are discussed, Closeness Centrality and Betweenness Centrality analysis to address this issue.

3.3.2 Closeness and Betweenness Centrality

Closeness centrality methods highlight the "*distance*" of a concept to all other concepts, and help to form the clusters within the network and clarify the connections between them. Betweenness centrality, on the other hand, helps to reveal the whole network structure and quantifies the number of times a network member represents a *bridge* that connects other concepts.

3.3.2.1 Closeness Centrality Analysis

In closeness centrality, the analysis takes clusters as units of examination in order to calculate how "*far*" or "*close*" they are from each other or from nodes outside the cluster. From information and influence spread perspectives, closeness centrality can be seen as a tool to calculate the time needed for information to spread to other members of the network sequentially (Newman, 2005). In this analysis, the network member can be regarded as an individual cluster, which means that this analysis can be applied to the binary connections as the individual concepts within the ontology.

The number of relationships between two clusters of concepts can be regarded as the related *power* between these clusters. This *power* can be seen as an indicator or a measure of the closeness between the clusters. Moreover, the number of connections offers a quantitative value that can be transferred to a value between zero and one. This value could denote the closeness between the clusters of concepts within the network.

For the purposes of this research, the closeness centrality is concerned with the significance of a seeding keyword S in producing a concept W, and the degree of control S had in the generation of W in the targeted domain. Actually, degree centrality function $f_{Cen}(t)$ tracks all the appearances of the term t in all the generated sets C irrespective of the starting keywords that generated t. Therefore, when a starting keywords S appears in the generated set C, we can define a function $f_{Cen}(t,S)$ that can calculate the occurrence of t in all the sets where S was a seeding word:

$$f_{Cen}(t,S) = \sum_{i,j} f_E(t, C_{(s_{pi}, s_{pj})}) \quad | \quad 1 \le i \le j \le n$$

Equation 3-7

Then, this is used to calculate the closeness distance $f_{Cl}(t,S)$ that represents the power of (S) in predicting t:

$$f_{Cl}(t,S) = \frac{f_{Cen}(t,S)}{f_{Cen}(t)}$$

Equation 3-8

It can be assumed that both functions $f_{Cl}(t,S)$ and $f_{Cl}(S,t)$ are not going to be identical, since both parts of the fraction would be different in both functions. Function $f_{Cen}(t,S)$ depends on the significance of the semantic relation between t and S.

The output from the diabetes case study discussed in chapter 5, show that relationships are directional. This gives applications the capability to examine and output the required ontology from various angles and viewpoints. In addition, low closeness values have highlighted possible relationships between conceptual clusters, and can play the role of bridging different domain concepts.

In order to study these potential bridge concepts and grasp the overall ontology relationship mapping, the next analysis - Betweenness centrality - is applied.

3.3.2.1.1 Betweenness Centrality Analysis

Betweenness centrality is a tool used for identifying bridging nodes within the network. It highlights and quantifies the importance of such nodes, which may have been ignored by the previous centrality analyses (Degree and Closeness) (Freeman, 1977).

The Betweenness centrality analysis is essential in discovering those nodes (or group of nodes) that exist in intersecting clusters, so the connections between various conceptual clusters or sections are clarified. For the purpose of this research, we use the Betweenness function $f_B(\mathbf{t}, \mathbf{S})$ where the nodes that have meaningful Betweenness values can be identified as follows:

- Using the closeness centrality computed via the function $f_{cl}(\mathbf{t}, \mathbf{S})$ discussed in the section 3.3.2.1, the nodes with small closeness centrality values in the ontology, i.e. their centred clusters are not close to each other compared with the remaining clusters are addressed. For example, the examination of $f_{cl}(\mathbf{t}, \mathbf{S})$ values may refer to two concepts w1, w2 where:

$$\boxed{f_{cl}(\mathbf{w}_1, \mathbf{w}_2) \to 0 \quad \textbf{and} \quad f_{cl}(\mathbf{w}_2, \mathbf{w}_1) \to 0}$$

Equation 3-9

- Betweenness analysis considers those nodes that are remotely located in either direction of w1 and w2. There may be a concept B that is well connected to both clusters:

$$f_{cl}(\mathrm{B}, \mathrm{w}_1) \rightarrow \max_{1 \le p \le i \le n} f_{cl}(\mathrm{B}, \mathrm{B}_{pi})$$
$$\text{and} \quad f_{cl}(\mathrm{B}, \mathrm{w}_2) \rightarrow \max_{1 \le p \le i \le n} f_{cl}(\mathrm{B}, \mathrm{B}_{pi})$$

Equation 3-10

The node B has bridged the gap between w1-centred cluster and w2-centred cluster. The presence of such a node in the ontology demonstrates that connecting concepts can be detected. It also shows that the outer concepts in the ontology network should not be ignored, as they could play an important role in bridging clusters in the domain ontology.

Both closeness centrality and Betweenness centrality measures have provided the ontology network with the relation measurement between various terms. This combined with the degree centrality analysis that provided the positional understanding of the concepts, allows the corpus network to operate and behave as ontology.

3.4 Domain Ontology Generation Summary

The knowledge source for the application of this research uses the OneLook Reverse Dictionary as the source for semantically related concepts. The generated concepts are linked with the OneLook knowledgebase to enhance the suitability of the derived ontology to text mining applications. This research applied Social Network Analysis techniques to carry out the network analysis of the ontological structure to create a richer and more useful ontology mapping. The generated ontology can be used to semantically annotate the text extracted from the online discussion forums.

Chapter 4: Semantic Text Analysis and Visualisation of Forums

Chapter three has described the techniques and tools developed to produce a domain ontology for the process of analysing the text in online discussion forums. In this chapter, we describe the semantic text mining method that incorporate the domain ontology and a suitable external general knowledgebase in order to annotate and analyse the text extracted from the diabetes discussion forum. The proposed system contains three main phases:

- The first is the semantic annotation and topic identification component.
- The second is a dynamic tag cloud generator.
- The co-occurrence network generator.

4.1 Semantic Annotation and Topic identification of Discussion Forums

As discussed in chapter two, this research proposes an ontology for tagging text in posts and it is specially designed to address the lightly (or non-existent) tagged text in the online discussion forums, where manually-added user tags are either rare or non-existent. This domain ontology is focused around the diabetes area. However, although the discussions in the forum cover topics that are closely associated with the diabetes domain, a significant part of the discussions is in the general yet related areas around diabetes. Therefore, it is important for the analysis of these discussions to be able to identify the topics that cannot be identified using the domain ontology alone.

This research proposes a new approach that improves the semantic annotation and topics identification processes. In the proposed approach, a combination of domain ontology and a general-purpose knowledgebase (DBpedia) are used to semantically annotate the posts in the forum and identify the topics in the discussion. This section presents the three main steps of a topic identification process that semantically chooses the topics discussed in the diabetes forum. Figure 4-1 shows the steps of the process:

1) Semantic annotation and Enrichment,

2) Subject (Topic) Identification, and

3) Topic Ranking and Selection.

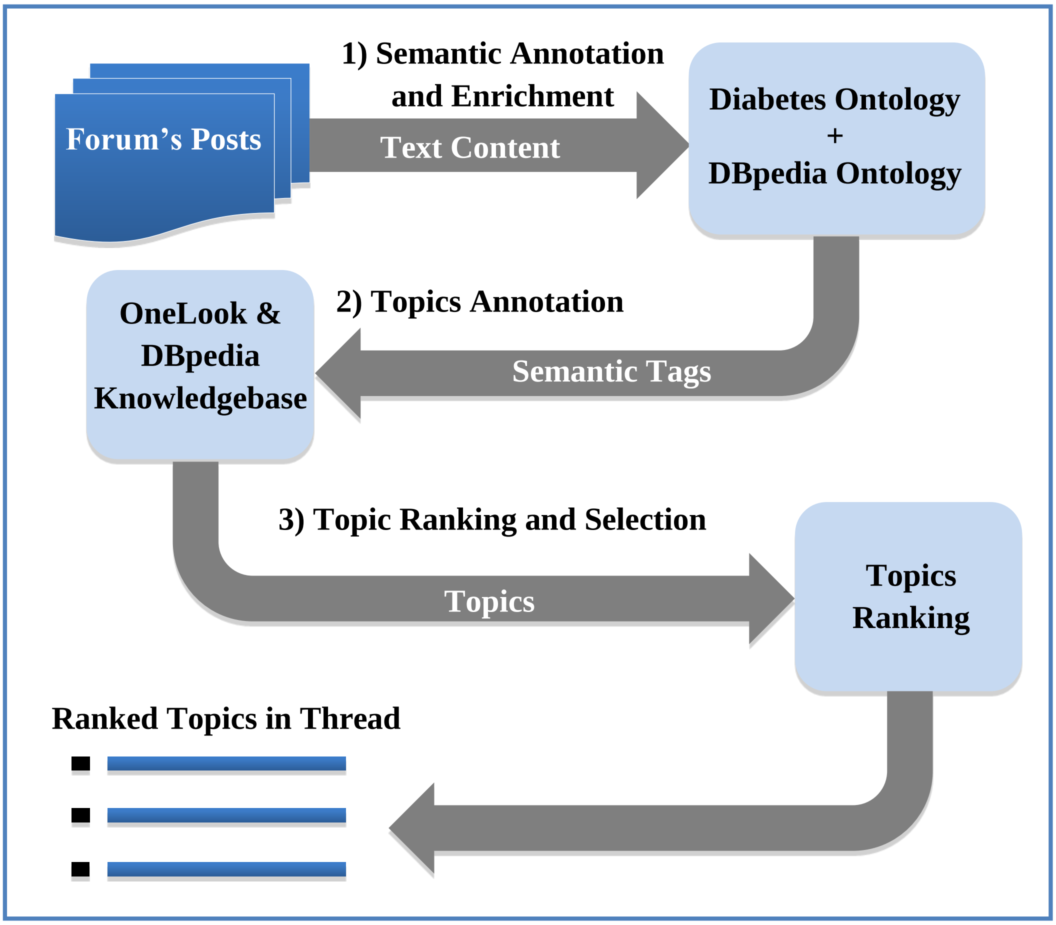


Figure 4-1: Topic identification process

4.1.1 Semantic Annotation and Enrichment

This step chooses the tags to annotate the text according to two criteria: The first is the degree centrality in the ontology (discussed in section 3.3.1) and the second is based on term (entity) extraction from a general knowledgebase. Hence, concepts that are more central to the domain and frequently discussed topics are ranked higher compared to terms that are not frequently discussed, nor central to the domain. The ontology used to improve the tagging process (which was developed via the process discussed in chapter three) is a domain-based ontology. Therefore, the proposed approach can be utilised with any text content as long as the domain ontology has been generated using the methodology discussed in chapter three.

4.1.1.1 Entity Extraction

The entity extraction step is inspired by the previous classical methods for Named Entity (NE) recognition, which rank important concepts based on their features. In the proposed approach, concepts that are more central in the domain are expected to be better candidates for tagging posts than other terms with less statistical significance in the text, which are not usually used as tags in the domain. Statistical significance methods do not take into consideration the semantic relationships between terms. As a result, they are not able to recognise similar or related concepts. This is the reason why most traditional Named Entity (NE) techniques proposed tags that do not represent what the users are expecting or seeking.

For instance, examining the terms that appear in a “Diabetes” tag cloud for one of the forums threads (Figure 4-2) that was constructed based on statistical significance criteria, some of the tags that represent aspects of the discussed topic enable users to focus on the related concepts such as “Insulin”, “diabetes”, or “diagnosed”. However, several tags, such as “jun”, “pm” and “bg”, do not offer any value to the text analyst.

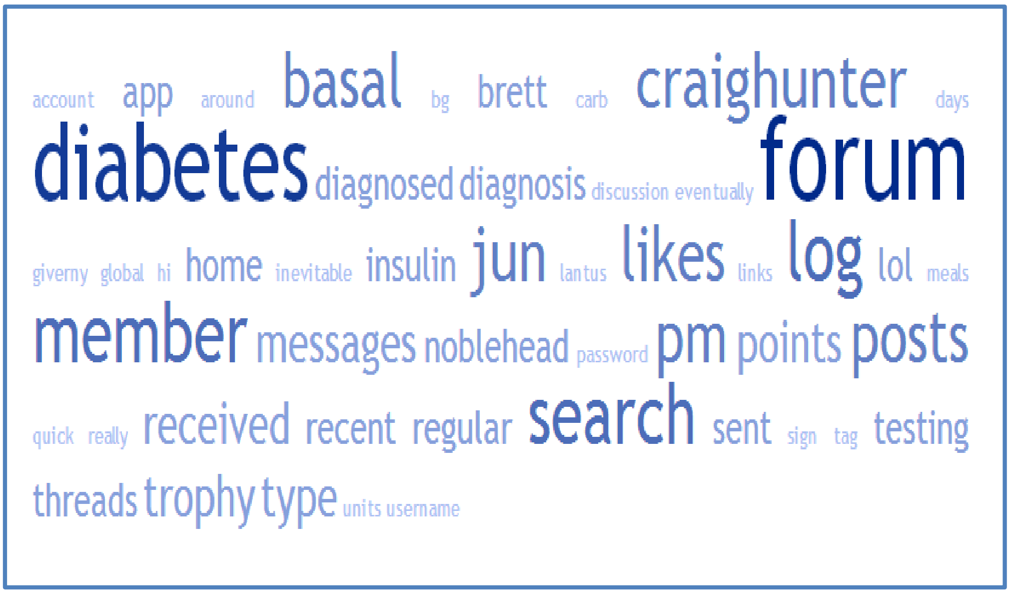


Figure 4-2: Tag cloud of posts from a diabetes forum thread generated by the TagCrowd application

These tags are considered important from the statistical point of view, but they are not part of the tag cloud that appears to have any human value. They are unlikely to be used to tag the text by human annotators. Our proposed approach can recognise and filter such tags in order to prioritise the use of tags that are more appropriate to tag the content and consequently are more likely to be helpful to the tag cloud users.

Unlike producing an annotation based on statistical techniques, the new system is based on a combination of domain ontology-based tagging and term extraction from the thread content using a general knowledgebase. Terms extraction from a given content in order to produce tag clouds is connected to the problem of "cluster labelling" (Manning et al., 2008). There are two general approaches to address this problem:

1) Statistical extraction of important tags from the content and

2) Using external resources such as folksonomies or online encyclopaedias.

Statistically extracted entities are not always considered ideal for annotation and tagging. Carmel et al. (2009) has illustrated that statistically significant terms extracted directly from the text do not agree with manual tags annotated by humans. In a number of cases, they found that statistical methods have struggled to recognise these manual concepts as worthy labels to the annotated text.

In our proposed system, while annotating the forums' posts, if the post text was semantically annotated with none or only a few concepts from the domain ontology, then the tagging process can be improved using extracted concepts from the text using an external knowledgebase. The blend of semantic tags and a concept extraction strategy is constructed through creating a sorted list of tags associated with each post, where the ontology tags are put at the top and the extracted entities are ranked under them in the list. The assumption behind this strategy is that ontology tags are more central to the domain compared to the extracted entities from the text.

An external resource such as DBpedia might be a solution to enhance the tagging process as it has a supply of semantic information that is kept updated frequently. In fact, DBpedia, as an RDF dataset that contains structured information extracted from Wikipedia, provides the opportunity to utilise its terms to semantically annotate web sources and recognise entities.

In this approach, terms can be linked either with their corresponding domain ontology concepts or with their DBpedia related sources. After this linking (mapping) step, explicit semantics could be associated to the terms and their "meaning" could be enhanced through the utilisation of the ontological structure of both the domain-specific ontology as well as DBpedia.

In order to get the full benefits from DBPedia and its structure, the semantic annotation process uses the DBpedia Spotlight tool (Figure 4-3), that utilises DBpedia to identify which terms in the text should be tagged and semantically annotated with DBpedia URIs.

Figure 4-3: DBpedia Spotlight Web Interface

Using Spotlight, the semantic annotation process could be controlled or constrained to specific categories/classes (or a group of categories/classes). Spotlight, can help to apply boundaries and restrictions on any section of the DBpedia ontology. In addition, we can configure Spotlight confidence parameter along with the requirements of the required tasks.

An example of using DBpedia to annotate a post from the diabetes forums can be seen in Figure 4-4, where the entities in the text have been matched with their corresponding URIs resources in DBpedia. Each URI resource includes the "Surface Form" which is the annotated text from the post, the "URI" which is the link to the DBpedia resource page, and "Support" which represents the popularity of the Wikipedia article associated with the annotation.

DBpedia spotlight API is used to annotate the posts extracted from the online forum. The resulting URIs are stored in a local database for further analysis.


```
.
.
.
<Resource
types="DBpedia:Plant,DBpedia:Eukaryote,DBpedia:Species,Freebase:/food/ingredie
nt,Freebase:/food" support="940"
percentageOfSecondRank="0.21825067497580442"
similarityScore="0.22970135509967804" offset="119" surfaceForm="orange"
URI="http://dbpedia.org/resource/Orange_%28fruit%29"/>

<Resource
types="Freebase:/book/book_subject,Freebase:/book,Freebase:/chemistry/chemic
al_classification,Freebase:/chemistry,Freebase:/food/ingredient,Freebase:/food,Fre
ebase:/food/food,Freebase:/food/nutrient,DBpedia:TopicalConcept"
support="2568" percentageOfSecondRank="0.4118355173206223"
similarityScore="0.17656682431697845" offset="229" surfaceForm="sugar"
URI="http://dbpedia.org/resource/Sugar"/><Resource types="" support="107"
percentageOfSecondRank="-1.0" similarityScore="0.15688271820545197"
offset="240" surfaceForm="sweetener"
URI="http://dbpedia.org/resource/Sugar_substitute"/>
.
.
.
```


Figure 4-4: An example of DBpedia Spotlight web service output

4.1.2 Subjects (Topics) Annotation

Having annotated and enriched the posts using the approach introduced in the previous step, the relationships between the semantic annotation tags are explored in order to find the topics of the posts.

For each annotation in the text, there is rich information available in the derived ontology and DBpedia in the form of properties that can be used to identify the topics.

A topic in this research case is a category, which is a value (or a number of values) that is assigned to the properties of each entity. For example, diabetes drug “Lantus” is tagged with “Insulin glargine” annotation. “Insulin glargine” is then used to get the subject associated with Lantus, which are “category:Insulin therapies” and “category:Peptide_hormones”.

In order to retrieve the topics of posts, first DBpedia is queried using the SPARQL language, as it allows for navigation in the DBpedia hierarchy to obtain the relationships between the semantic annotations and their topics. Such relationships are represented in the DBpedia ontology via the *dcterms:subject* property that is assigned to the concepts, and *skos:broader* property that is assigned to categories. In the “Lantus” example, the topic identification brings back “dbpedia-owl:Drug” as the topic associated with “Lantus”.

However, a concept or an entity can be located in the different levels of the DBpedia hierarchy and therefore the proposed approach would result in obtaining several topics at different levels. This extracted information about the topics and categories is useful only if it was ranked in order to find the most related and representative topics of the posts.

4.1.3 Topic Ranking

In this step, the most representative subjects of the posts are selected. In order to achieve this goal, the topics within a particular thread in the forum are sorted based on their frequency and similarity. As a result, a ranked list of topics is generated for each thread in the forum.

The output of the semantic annotation system is annotated posts with semantic tags associated with them. These tags can be used for graph-based analysis of the forum and its discussions. Techniques and methods for visualising information are able to reveal hidden information about the patterns in a data set and the relationships between different data elements. This provides the analysts with a powerful tool to explore and interact with the data. Therefore, information visualisation can be utilised as an effective method for text representation and analysis.

The two main methods for analysing the relationships between these semantic tags and uncover hidden trends are tag clouds (will be discussed in section 4.2) and co-occurrence network (will be discussed in section 4.3).

4.2 Tag Cloud Visualisation

Tag clouds are one of most commonly used visualisations on the World Wide Web. They provide the user with a compacted representation of the text in an individual webpage or a website. This representation uses a group of words (tags) whose frequency is reflected through the size of the words or a combination of size and colour (Viégas and Wattenberg, 2008). Some researchers, such as Bumgardner (2006), have considered tag clouds as a special type of weighted word lists, and hence visualised them as one. Figure 4-2 shows a traditional tag cloud, where the size of the tag reflects its importance.

Over the past decade, tag clouds have been used in websites and applications for various purposes, such as visualising free form text, or for visual data browsing and queries. For instance, tag clouds have been utilised to enhance the browsing function of social media, such as blogs. An example of using tag clouds in social media is the study by Phelan et al. (2009), where they introduced a twitter-based recommendation system in order to recommend related news articles. Figure 4-5 illustrates this system.

The tag cloud was used to explain to the user the concept space from which the news results were derived.

Figure 4-5: Using tag clouds in Twitter-based recommendation system, Source: (Phelan et al., 2009)

Hassan-Montero and Herrero-Solana (2006) used word clustering to re-arrange words and group similar tags, as well as the re-grouping of the lines in similar clusters. Figure 4-6 shows an example of a clustered tag cloud, where each line includes tags that belong to a specific cluster.

Figure 4-6: Clustered tag cloud, Source: (Hassan-Montero and Herrero-Solana, 2006)

4.2.1 Tag Cloud Advantages and Drawbacks

Several researchers have argued that the use of tag clouds can have many positive influences on basic visual tasks and text analysis due to:

- Its compact layout;
- Its capability to show different dimensions at once, such as size, which allows users to spot the most frequent terms quickly within tag clouds, or colour that can be assigned to different clusters in the tag cloud.

Hence, tag clouds are easily "scannable" and provide reasonable overview of the content. When compared with the plain text (where researchers such as Weinreich et al. (2008) has shown that average user reads around 20% of the web page text), these advantages of using tag clouds are useful. In addition, Rivadeneira et al. (2007) showed that tag clouds support various user tasks such as browsing, searching and providing a good outline and impression of the underlying content.

Although tag clouds have advantages, there are a few drawbacks to using them. Researchers, such as Hearst and Rosner (2008), have shown that with the traditional static tag cloud design, longer words get more user attention compared with the shorter words. In addition, they found that visual closeness of tags in the traditional cloud is meaningless and hence the important relations and associations cannot be inferred, and visual comparisons are complex and challenging.

However, many of these disadvantages can be addressed by adding improvements to the tag cloud visualisation regarding the positioning and sorting of the tags, viewing considerations, or integrating tags relationships.

There are several methods to map visual features to the data set, which could be utilised to enhance tag clouds. Bumgardner (2006) stated that visualisation clarity requires that the mapping of tag cloud features should be meaningful and proposed the use of interesting associations, such as (Font size ↔ time) that shows the more recent words bigger than older words. Another example, is (Font type ↔ decade/century), so older data has old font types. These improved tag clouds promise a better visual attraction and easier "scannability" by users. Other studies used "Stacked Graphs" for visualising tags. These graphs are helpful techniques for visually analysing the differences and similarities of a set of words over time (Byron and Wattenberg, 2008; Havre et al., 2002).

Tag clouds can be classified into two types from time evolution perspective: Static and Dynamic visualisation (Cui et al., 2010). The previous examples of tag clouds were mostly static tag clouds. The techniques for producing a static tag cloud concentrate on addressing common issues (such as overlapping of tags) in order to enhance the general understanding of the cloud, while the dynamic visualisation demonstrates the evolution of content over time in several documents or articles.Next follows a discussion and a comparison of the two types in more detail.

4.2.2 Static Tag Clouds

The most commonly used static tag cloud is a rectangle consists of sequential lines of words (tags) arranged alphabetically in a stacked lines design (Figure 4-2).

The static nature of these clouds is the main disadvantage of using them with social media streams. The main challenge in visualising social media streams lies in providing a suitable high-level overview that can capture the evolution of the content and topics over time. Therefore, dynamic tag clouds are needed in order to address this issue.

4.2.3 Dynamic Tag Cloud

Researchers proposed different methods to visualise the tags development over time. For example, Dubinko et al. (2007) demonstrated a dynamic tag cloud for Flickr (www.flickr.com), with which the users of Flickr can spot and observe the tags as they change over a period of time (Figure 4-7). An animation in a web browser allows the user to detect and interact with the tags as they evolve over time.

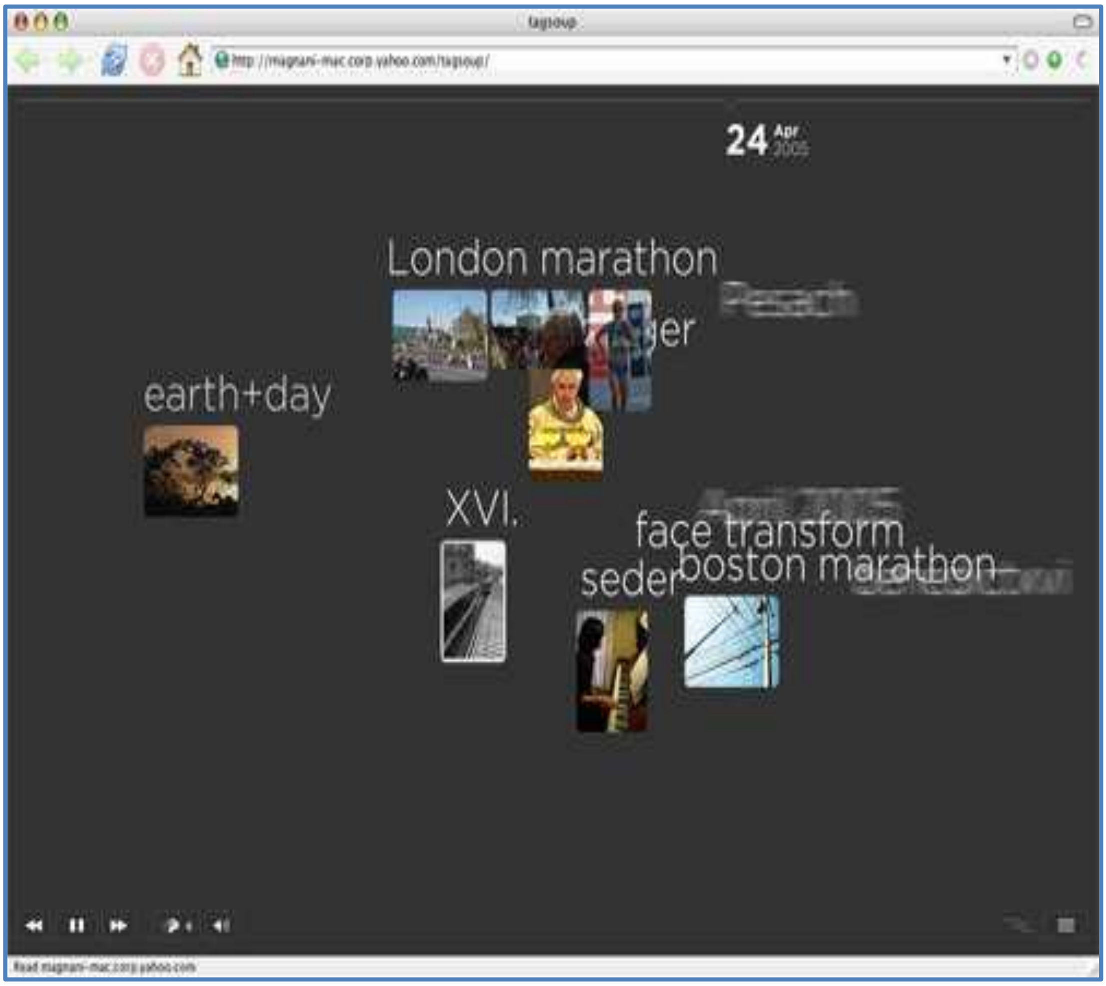


Figure 4-7: A dynamic tag cloud for Flickr, Source: (Dubinko et al., 2007)

Another example is research by Russell (2006) which proposed a tool to visualise the development of clouds over time (Figure 4-8). The system takes a request for a URL, then obtain the tagging data from del.icio.us, and plots the activity of users tagging over time.

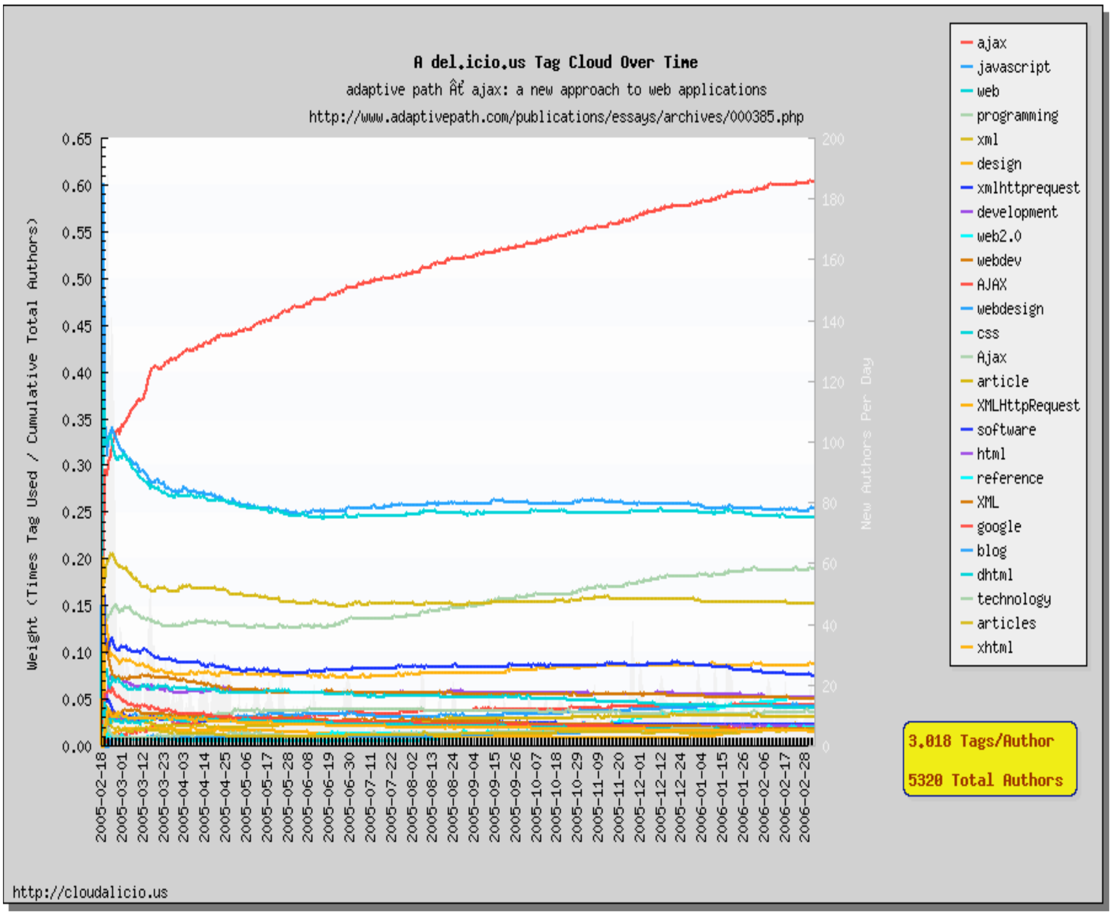


Figure 4-8: Development of tag cloud over time, Source: (Russell, 2006)

These approaches generate a graph that represents the evolution trend of the tags within the cloud instead of the visualisation of the evolution of the tag clouds themselves over time.

In order to address these issues, a new dynamic tag cloud model is suggested. This model would compare and visualise the clouds and their content changes over time.

4.2.4 Diabetes Forums Semantic Tag Cloud

The methods and technique for producing the dynamic tag cloud start with extracting the tags and topics from the content of the text, then using the output to produce the cloud. This process was described in section 4.1. This section will describe and discuss the proposed enhanced technique for the tag cloud generation.

4.2.4.1 Tag Cloud Generation

The proposed system includes two major elements:

1) An Importance curve, and

2) Time-based tag clouds.

The chart window portrays the changing importance of the forum content over a period of time, modelled by a group of central demonstrative topics/tags. The line chart in the chart window displays the varied importance of the time-based tag clouds obtained from analysing and tagging the forums posts (illustrated in section 4.1). The horizontal axis of the chart represents the time, while the vertical axis represents the importance of the tag clouds. Time-based tag clouds provide an illustration of the changing topics and content over time within the forum. In order to improve the readability and facilitate the analysis of the tag cloud, different **colours** are used to reflect the tags behaviour over time:

- **First-time Appearance**: If these tags appear in the succeeding tag clouds, then they are coloured in Yellow, otherwise, they are coloured in Red (unique appearance);
- **Last-time Appearance**: If these tags appeared in preceding tag clouds, then they are coloured in Green, otherwise they are coloured in Red (unique appearance);
- **In-between Appearance**: These tags appear in both earlier and later tag clouds, and they are coloured in black.

The concept behind this visualisation is that text content of the forums and its corresponding tags evolve and vary over time. Therefore, the visualisation of this type of data is unique but difficult to implement at the same time due to the dynamic characteristic of the data. The author suggests a solution adapted from Wang et al. (2008)'s significance-driven approach in order to visualise time-varying data based on analysing the spatio-temporal behaviour of the data using the concept of conditional entropy from information theory. The argument is that volume data at a specific point in time is more important when it contains more unique information, i.e. little or no information overlaps with volume data at other points of time. Therefore, the significance of the data at various points could be measured and calculated quantitatively using information theory entropy measures.

Using the concept of importance analysis for volume data visualisation, this research proposes an adaptation for this analysis to be used for tag clouds in order to illustrate the varying importance of a forums content over time. The tags inside the tag clouds have a dynamic behaviour that can be studied and analysed in order to effectively visualise and illustrate the nature of this dynamic behaviour to researchers and tag clouds' users. Therefore, the importance of a tag cloud at a specific time point depends on: i) the information it contains and ii) the overlapping information with other tag clouds.

In order to explain the proposed visualisation, a few concepts from information theory will be introduced: Shannon Entropy, Mutual Information and Conditional Entropy.

I. **Shannon Entropy (Information Theory)**: The concept of entropy in information theory is related to the information content, and it measures the *unpredictability* or the *uncertainty* of a random variable (Cover and Thomas, 2006). It can be calculated using the following formula:

$$H(X) = E[I(X)] = E[-\ln(P(X))]$$

Equation 4-1

Where *H* (Eta in Greek) is the Shannon entropy value of the random variable *X* with values in ($x_1,\ldots, x_n$), *E* is the average operator (expected value), and *I(X)* is the information content of the variable *X*.

When entropy is taken for a finite sample, it can be calculated as follows:

$$H(X) = \sum_i P(x_i) I(x_i) = -\sum_i P(x_i) \log P(x_i)$$

Equation 4-2

Where P(x) is the probability mass function (PMF) of X.

The entropy is maximised when the variable has maximum uncertainty. For example, when tossing a coin, the entropy of the next toss equals to one (when the coin is fair, i.e. the probability =1/2) since this situation is the hardest to predict the result of the next coin toss. Whereas the entropy equals zero when there is no unpredictability. In the previous example, the entropy equals zero if the two sides of the coin are identical (both are heads or both are tails).

II. **The mutual information measure:** This concept assesses the *mutual dependence* between two random variables (Cover and Thomas, 2006). Simply put, this calculates the information that the random variables share, i.e. it calculates how knowing one of the variables affects the certainty of the other. For instance, if two random variables X and Y are independent variables, then the mutual information equals to zero. This means that knowing one of them would not reduce the uncertainty of the other and vice versa. Another example is when either one is a function of the other (for example, $Y = f(X)$), then knowing the value of the variable X will determine the value of the variable Y. The mutual information value equals to the uncertainty of Y in this case, i.e. its entropy value. The mutual information function can be defined as follows (Yao, 2003):

$$\boxed{I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right)}$$

Equation 4-3

Where p(x, y) is the "*joint probability mass function*" of X and Y.

It can be seen from the formula that mutual information is a symmetric measure, i.e., I(X; Y) = I(Y; X), and that I(X;Y) ≥0.

III. Conditional Entropy: Having defined the mutual information, the conditional entropy of the random variables X and Y can be calculated as:

$$\boxed{H(\mathrm{X} \mid \mathrm{Y}) = \mathrm{H}(\mathrm{X}) - I(X;Y)}$$

Equation 4-4

Naturally, if the entropy H(X) is considered a method to gauge the uncertainty of X, then the conditional entropy H(X|Y) can be regarded as a measure of the remaining uncertainty of X when the random variable Y is known. Therefore, in order to evaluate the significance of a tag cloud X, the entropy H(X) is calculated, then the mutual information I(X;Y) is calculated in the presence of the tag cloud Y. These two values are then used to measure the significance of X through H(X|Y). In the proposed visualisation, the entropy of the tag cloud X is calculated compared with other tag clouds Y, and then the importance of all tag clouds is computed to form the *Importance Curve*.

4.2.4.2 Tag Cloud Importance

As discussed in 4.2.4.1, the tag cloud is deemed important if it has high information value. This is partly through not sharing much information with other tag clouds. This is analogous to managers monitoring deviations away from the norm as a better signal of significance for driving subsequent action. The tag cloud importance value can be calculated using the conditional entropy concept previously described. Calculating this value for tag clouds over time yields a value vector. Drawing this vector in a two-dimensional chart (time & significance) results in what is called an *Importance Curve* (Wang et al., 2008) which illustrates how the significance value of a tag cloud varies over a period of time. In order to calculate the importance of a tag cloud, the entropy of that cloud $H_t(C)$ at time t is evaluated. Next, the mutual information function $I(C;Y)$ is calculated between the tag cloud C and its preceding and succeeding tag clouds. Then using the conditional entropy formula (Equation 4-4), the importance of the tag cloud can be derived.

In order to estimate the tag cloud entropy H(C), each tag in the cloud is first represented using a vector that depict the tag features. In this research, this vector contains the tag frequency (which is represented by the size of the tag) and the tag colour (described in 4.2.4.1). Then, a histogram (Figure 4-9) is generated using these tag vectors, where each column in the histogram represents the number of volume elements in the tag cloud that belong to specific combination of values. This histogram is used to estimate the entropy H(C) of the tag cloud C utilising its values as the probability in the entropy formula (Equation 4-2).

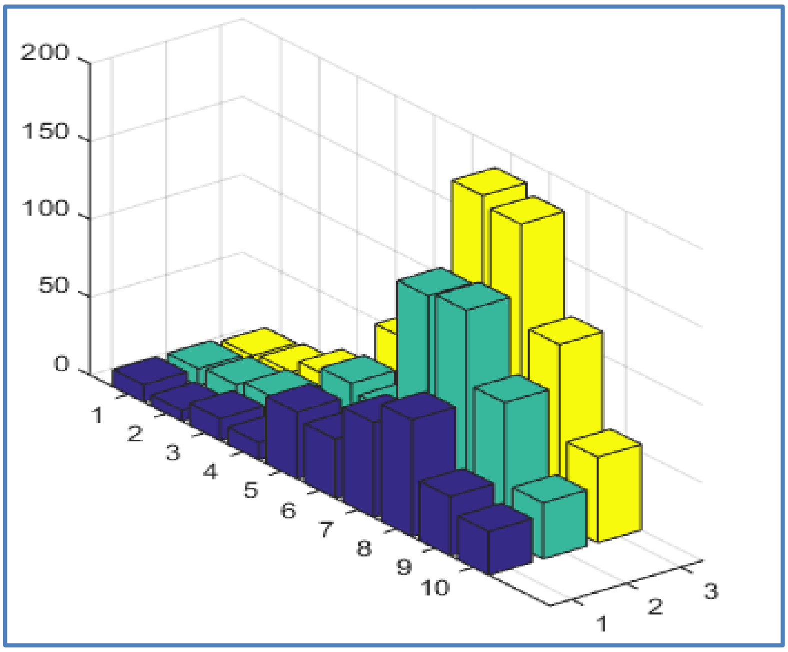


Figure 4-9: Histograms for estimating entropy of a tag cloud

There is a difference that the vector which represents the tags is used instead of the variable value. In this case, the entropy of a two-dimensional variable X is:

$$H(X) = -\sum_{a \in A} \sum_{b \in B} P(a,b) \log P(a,b)$$

Equation 4-5

Where A and B are X's elements in the two dimensions respectively.

Having computed the probability p(x) for each tag cloud and its entropy, then there is a need to calculate the joint probability p(x, y) in order to calculate the mutual information value I(X; Y) based on Equation 4-3. Therefore, a *joint-histogram* (Figure 4-10) is built based on both tag clouds X and Y. In this case, the information shared between two tag clouds can be defined as the common tags between X and Y, while the rest of the tags are regarded as independent tags.

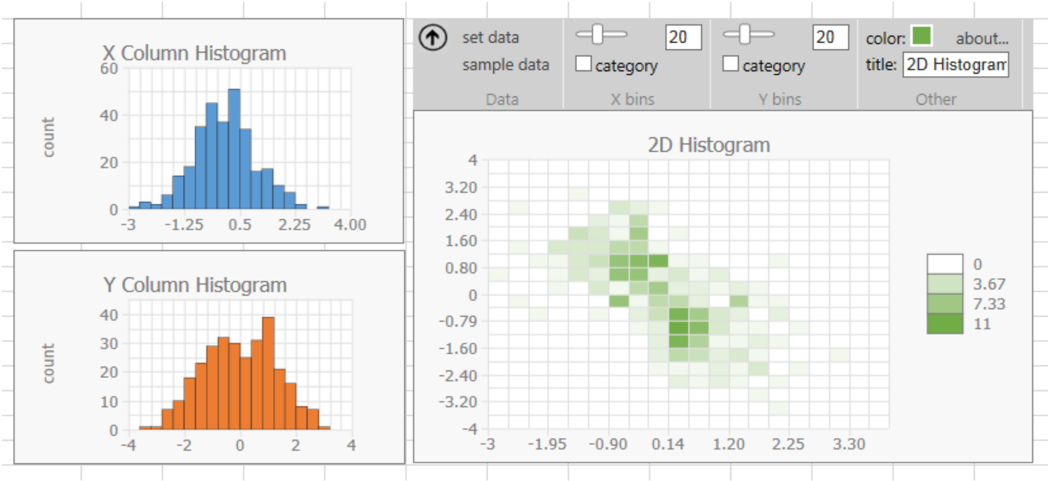


Figure 4-10: Two-dimensional joint-histogram

Hence, the resulting joint histogram is built through the number of the tags that belong to a specific histogram interval. Consequently, the mutual information value for X&Y can be estimated using the histogram columns.

The remaining configuration is related to the tag cloud conditional entropy: for a particular tag cloud X at time *t*, related neighbouring tag clouds should be chosen in order to calculate the conditional entropy of X at the time t. For practical reasons, these neighbouring tag clouds are within a time window centred around X's time point, i.e., the chosen tag clouds are before or after X on the time axis within a specific period of time (Figure 4-11).

1	2	4	0	4	2	1

Figure 4-11: Example of a time window with weights given (window size 7)

Then, the importance of a tag cloud X at a time t can be defined using this formula:

$$Imp(X) = \sum_{i=1}^{S} w_i \,.\, (H(X) - I(X;Y_i)) \qquad \textbf{Equation 4-6}$$

Where S is the window size and (w_i) is the normalised weight associated with the tag cloud Y_i in the time window, i.e. $\sum_{i=1}^{S} w_i = 1$.

4.2.4.2.1 Importance Curve Chart

After the calculation of the conditional entropy of the tag clouds over the chosen period, an importance curve can be plotted, which shows the changing semantics of the text content over time (Figure 4-12). The user can interact with the chart using the Sliding Bar option under the importance chart, where they can select a particular tag cloud.

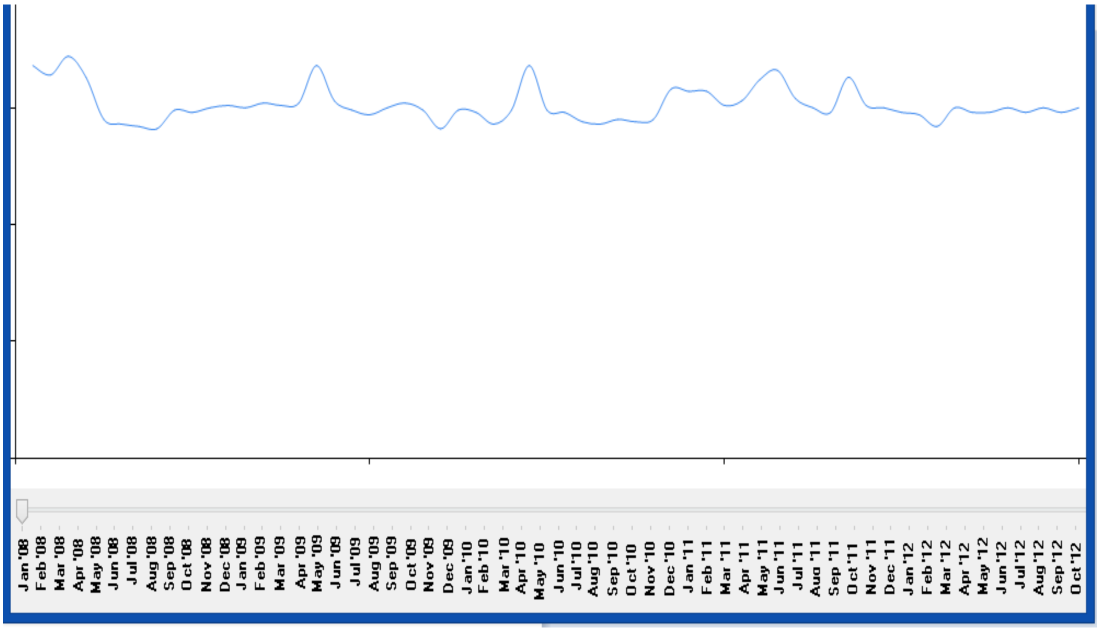


Figure 4-12: Importance Curve Chart

The data that web users in general, and social media users in particular, post on forums and review websites, provides an opportunity to have an insight into the domains of interest and their semantic network, which is reflected via the co-occurrence of tags and descriptions of concepts and tags in the written text.

4.3 Co-occurrence Network Analysis

Researchers like Saiz and Simonsohn (2008), have provided interesting evidence for the power and validity of utilising the relative frequency of discussed concepts in a phenomenon, in order to reveal the likelihood of a corresponding occurrence. However, going beyond just utilising the frequency of terms alone, it is valuable to assess the similarity/closeness among the concepts within the forum based on their co-existence in the discussion.

The idea of using occurrence frequency and co-occurrence as a representation of similarity between concepts in the text goes back to research in Knowledge Discovery (KD) literature. He (1999) reviewed the development of "co-word analysis", which is used to find and uncover connections between subjects in a particular research domain and therefore track the development of various topics and research interests in that domain as a whole, via tracing the frequency that two concepts co-occur in that domain.

A rich body of research in cognitive learning proposed that human beings establish network of associations that link between separated concepts, which form the stored knowledge (Anderson, 2005). Stuart and Hulme (2000) findings specify that associative connections between concepts in the long-term memory have significant effect on the recall performance of the short-term memory. The *Spread of Activation* theory proposed by Anderson and Pirolli (1984) suggests that when a node in the cognitive network is activated (e.g., Diabetes), it is likely that other associated nodes are activated in the same network, such as "Sugar" or "Insulin". The length of the connection between the nodes indicates the strength of the "Semantic relatedness" between the concepts (Herr et al., 1996). Hence, two closely associated concepts in the cognitive network are expected to be recalled from the memory and used simultaneously by the individual.

Özgür et al. (2008) used time stamped news articles to visualise the co-occurrence of people in Reuters newsfeeds. 21,578 news articles were sorted based on their dates and they were processed manually to identify people' names, which became the nodes of the network (Figure 4-13). Then, a link is built between two nodes if the corresponding names appeared in the same news article. This link has a weight that reflects the number of times the two have appeared together in the data set.

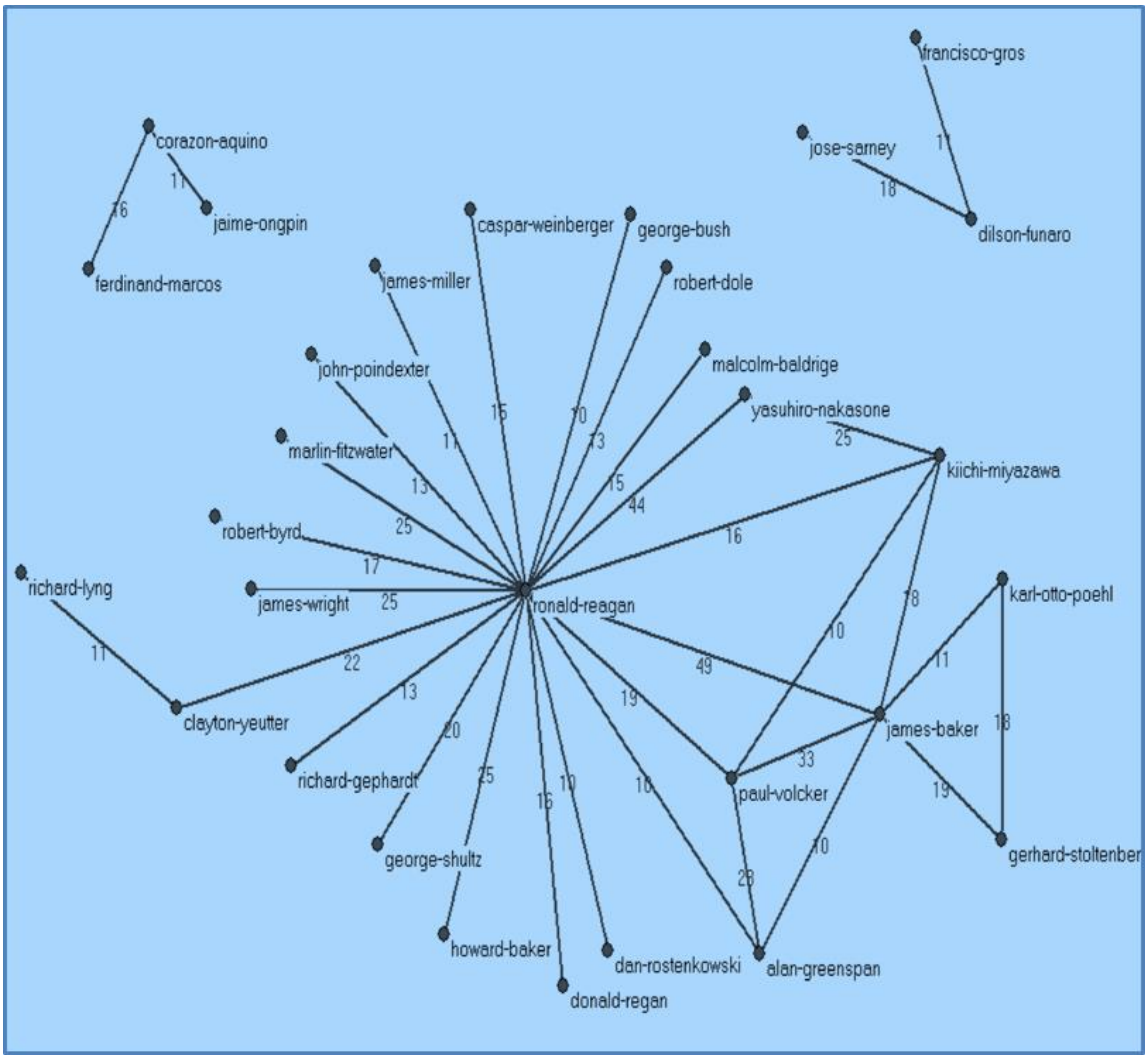


Figure 4-13: Strong ties analysis in co-occurrence network of Reuters' news, source: (Özgür et al., 2008)

In a similar way, the concept of co-occurrence was used in several approaches to build brand associative networks, like the study by John et al. (2006) and Teichert and Schöntag (2010), since product and brand attributes are more likely to be mentioned frequently with that brand in most customer reviews. Henderson et al. (1998) was the first to show the importance of consumer associative networks in order to recognise the connections between different brands and consumer perceptions, and map an array of branding effects such as the market structure, branded features, competitiveness, and segmentation. Collecting consumer opinions was carried out using a survey-based approach, since online forums and review sites were not commonly used at that time.

Netzer et al. (2012) used co-occurrence network to explore the reviews of the customers and obtain insights of the market structure automatically. Their approach utilises machine learning methods with both rule-based and dictionary-based mining techniques. They were able to extract products (cars and their models) using a training dataset that was tagged manually. Then, rule-based text mining was used to adjust the extracted nouns. These rules were specific to the considered domain (Cars Market) and were used also to sort out the concepts and perform names disambiguation. Using their approach, the companies can observe the market structure and their position at a low cost compared with traditional techniques of obtaining this information.

However, several issues limit the use of their system in other domains:

1. **Supervised Machine Learning (SML) and manually created rules**: Since each forum has its unique facets and usually specific domain aspects, the used approach needs to be adapted to suit the domain considered. The use of SML and rule-based mining required manual labour in order to tag and annotate the dataset. Other

methods to automatically tag the text are required to minimise the need for human intervention.

2. **Method generalisation**: The transferability of efforts from one domain to the other is an important challenge that surfaces in text mining methods. In the approach proposed by Netzer et al. (2012), the dataset used for training and the rules created are specific for analysing the car market. If the approach is to be used in another domain, new dictionaries are required for the new domain entities and concepts in order to help the expert prepare that dataset for training, although some parts of the automated process can be transferred, such as the POS tagger. In addition, the majority of the rules are also specific to the cars market and needs to be recreated for other domains. To sum up, parts of the process have limited transferability to other domains, while other parts are relatively easier to transfer.

Therefore, we propose a co-occurrence network model that utilises the semantic annotations and topics identified through analysing online discussion forums.

4.3.1 Diabetes Co-occurrence Network Configuration

A co-occurrence network of entities is produced where nodes are tags and the edges are the co-occurrence of these tags in the forum posts. The analysis of co-occurrence is helpful when looking for patterns in the forums' posts and building a semantic network from the analysed data. The steps to generate this network are:

1. All posts from the forums are processed and semantically annotated using the process described in section 4.1.
2. The nodes of the network are defined as distinct tags, and each node is associated with a weight, which is the tag frequency.

3. A relationship is constructed between two tags if they appear in the same post. This relationship carries the weight of the tags co-occurrence.

It should be noted that the output network is undirected in this case since the co-occurrence is symmetric, but it is a weighted network as the edges carry the co-occurrence value. This is different to the co-occurrence network proposed by Özgür et al. (2008) where there is a relationship between two connected people regardless of the number of co-occurrence. This means that the network produced in their case might be misleading when calculating the *closeness* of the nodes. This is also different from the associative network generated by Netzer et al. (2012) where all nodes are treated the same with no weighting (frequency).

In this research, using a simple occurrence measure to calculate similarity between concepts has one major drawback: Concepts with high-frequency appearance in the discussion text would have higher co-occurrence with almost all terms in the text compared with concepts that have low-frequency appearance. To address this, the plain occurrence of the each concept is normalised in the discussion forum using a method from information theory called "Pointwise Mutual Information" (PMI) (Church and Hanks, 1990). PMI is defined as the ratio between the actual co-existence of two concepts x&y to the probability of the co-existence between these concepts. The PMI between x&y can be measured as follows:

$$\boxed{PMI(w_1, w_2) = \frac{P(w_1, w_2)}{P(w_1) \times P(w_2)}}$$

Equation 4-7

Where:

- *P(w)* is the appearance probability of the concept w in a particular forum post.

- *P(w_1,w_2)* is the appearance probability of both concepts w_1 and w_2 in a particular post.

Pointwise Mutual Information (PMI) values give an indication of the relationships between the terms:

- PMI<1: This suggests that both concepts co-occur together less than expected by their separate appearances in the text.
- PMI>1: This suggests that both concepts co-occur together more than expected by their separate appearances in the text, which shows that this co-occurrence is more than just a statistical chance, which is an indicator of their semantic relatedness.

4.4 Summary

The semantic text mining method derived for analysing diabetes discussion forums was described in this chapter. Three main components are introduced:

1. A hybrid semantic annotation and topic identification component using a domain ontology and the DBpedia knowledgebase,
2. A dynamic Tag Cloud Generator based on principles from information theory, and
3. A temporal co-occurrence network generator.

The proposed text mining method needs to be evaluated through testing the process for the case study, diabetes online forums.

Chapter 5: Semantic Text Mining Results & Evaluation

In chapters three and four, methodology and techniques that can be used to create a new ontology network and then analyse online discussion forums were explored. The case study problem using the diabetes discussion forum described in chapter one is used in this chapter to assess the ontology derived by applying the methods described in chapter three and the enhanced semantic text mining process described in chapter four.

5.1 Diabetes Ontology Corpus

Three pairs of concepts were chosen as the seeding keywords for the ontology building from suggestions by a Diabetes expert from Warwick Medical School. These keywords were chosen to cover different aspects of Diabetes:

1. “Diabetes Mellitus” & “Insulin”,
2. “Diabetes Drugs” & “Diabetes Treatment”,
3. “Diabetic Diet” & “Diabetes Food”.

Using these seeding keywords as an input to the process explained in Section 3.2, “1622” concepts and “6431” relations between them were produced using the algorithm designed for the method (The raw data available in Appendix B). An ontological analysis of these results was carried out using the social network analysis techniques described in chapter three.

5.1.1 Degree Centrality Analysis of Diabetes Ontology Corpus

As discussed in section 3.3.1, the nodes of the ontology network that appeared in the results more frequently than other nodes can be considered as more “*central*” and more representative members of the domain.

Degree centrality analysis of the diabetes ontology network shows a significant variation of centrality values. Within this diabetes network of 1622 term members, degree centrality values range from high values such as "gestational diabetes" with 696 degree centrality, to low values such as 1 shared by 230 concept terms. The average degree centrality was 28.14 (45655/1622). The long tail part of the generated terms was a relatively low percentage at 14.18% (230/1622=14.8%, where 230 is the number of terms with centrality=1 and 1622 is the total number of concept terms).

The degree centrality distribution was of the same form as produced for the output of the configuration study experiments (section 3.3.1) as Figure 5-1 illustrates. The vertical axis in this figure represents the degree centrality value.

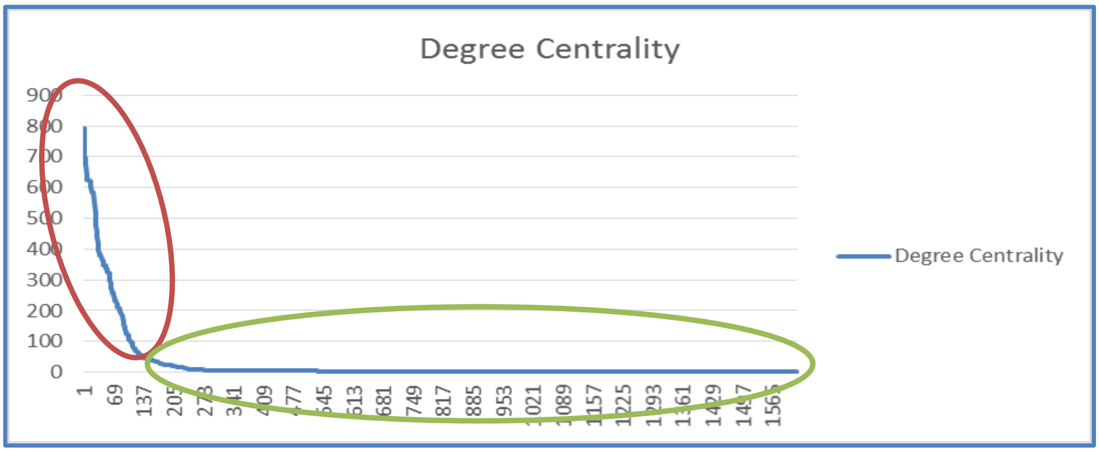


Figure 5-1: Degree Centrality (overall) in the diabetes ontology experiment

Figure 5-1 shows two clearly distinguishable sections: A highly centralised part highlighted in red ellipse and a long tail circled by the green ellipse. These two distinguishable sections highlight that not all the ontology terms are clustered within the same level of connection frequency. Therefore, they require further analysis (other than simple degree centrality) in order to identify the differences and similarities between terms centralities. Figure 5-1 has highlighted two different regions; Further analysis is split into analysis of these two parts in order to understand the changing nature of the centrality in greater detail.

5.1.1.1 Degree Centrality Increment analysis

Mathematically, the rapid drop part of the degree centrality curve in Figure 5-1 (red ellipse) indicates that the change rate in degree centrality values is higher than the growth of the keywords, i.e. $\Delta y/\Delta x \geq 1$. On the contrary, the long tail flat section in Figure 5-1 (green ellipse) indicates a relatively slow change rate of degree centrality values, i.e. $\Delta y/\Delta x \leq 1$. We can define a function called degree centrality increment that measures the change in degree centrality as follows:

$$\boxed{f_{DCI}(m,n) = \frac{f_{Cen}(\mathrm{k}_m) - f_{Cen}(\mathrm{k}_n)}{f_{Co}(\mathrm{k}_n) - f_{Co}(\mathrm{k}_m)} \mid \mathrm{k}_n, \mathrm{k}_m \in S} \qquad \textbf{Equation 5-1}$$

Where, $f_{Cen}(\mathrm{k}_m)$ is the centrality value of term k_m, and $f_{Co}(\mathrm{k}_m)$ is the term ID of k_m, and S is the set of terms in the ontology corpus. Figure 5-2 shows the degree centrality increment curve (Vertical axis represents the increment values).

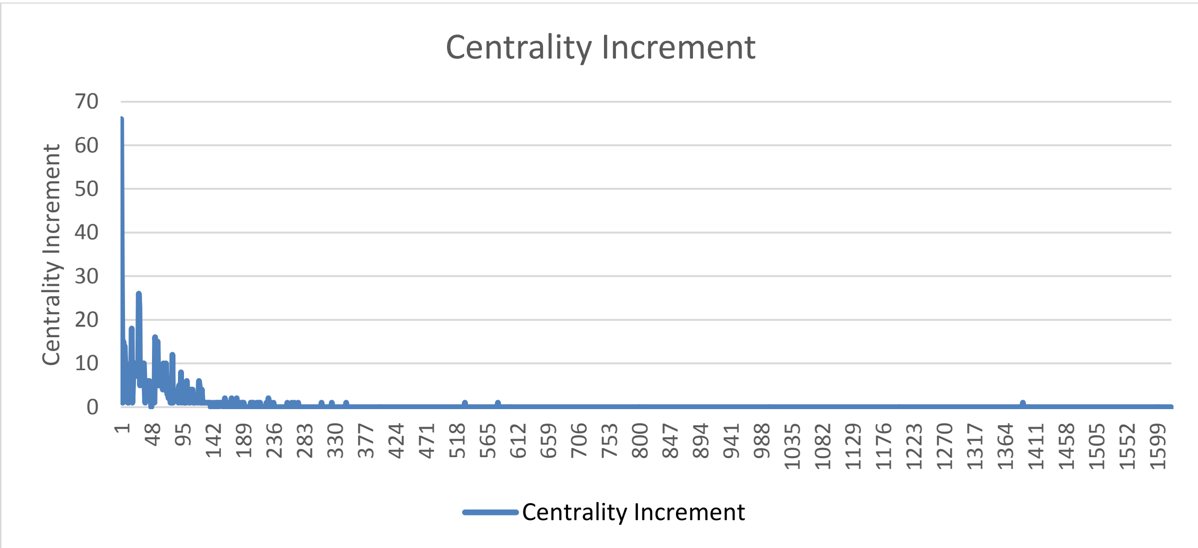


Figure 5-2: Diabetes ontology experiment – Degree Centrality Increment

In order to identify the points in the red ellipse section where the increment equals zero (cut-off points), we "zoom in" the first section of Figure 5-2. The result is shown in Figure 5-3, where the degree centrality increment for the first 140 terms in the ontology corpus are shown.

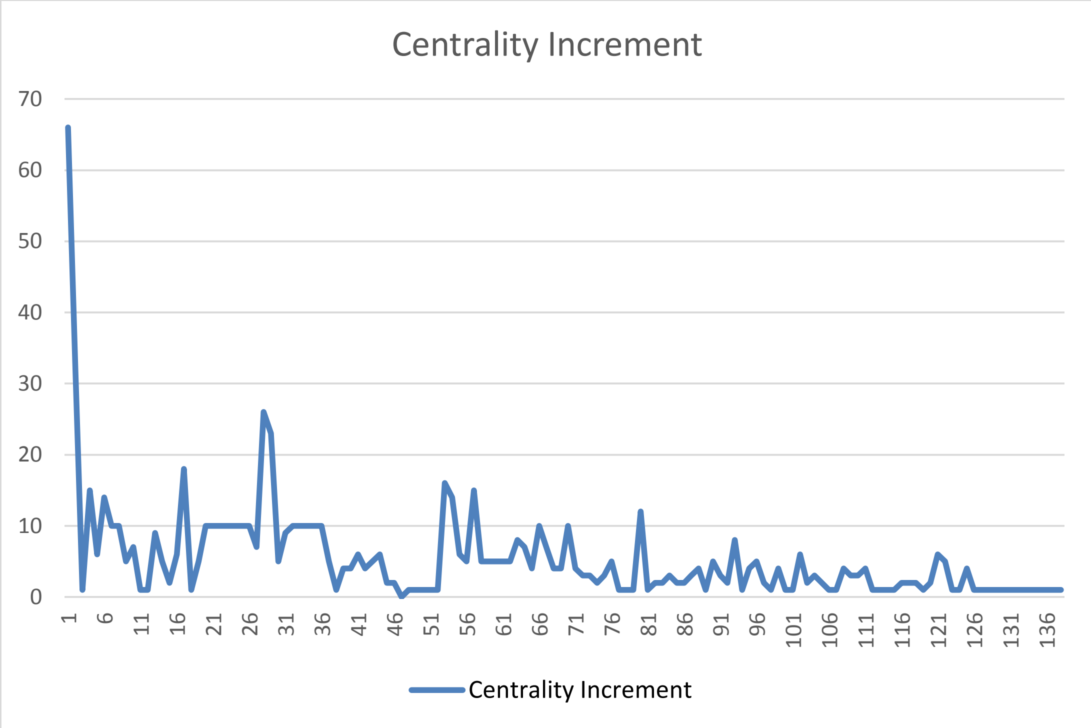


Figure 5-3: Diabetes ontology experiment – Degree Centrality Increment (First 140 terms)

The curve in Figure 5-3 cuts the horizontal axis (y=0) at point 47 (the corresponding diabetes term is "Melituria"). This highlights another section of the centrality distribution curve and therefore, the first section of the degree centrality values in Figure 5-1 (red ellipse) is actually composed of two main sections.

Thus this analysis reveals a tri-sectional curve in the diabetes ontology corpus through conducting a "degree centrality" and "degree centrality increment" analysis, which is similar to the tri-sectional curve observed by Ma et al. (2014). These three sections (zones) are shown in Figure 5-4:

1. Definition top zone;
2. Gradient change zone; and
3. Long-tail connection zone.

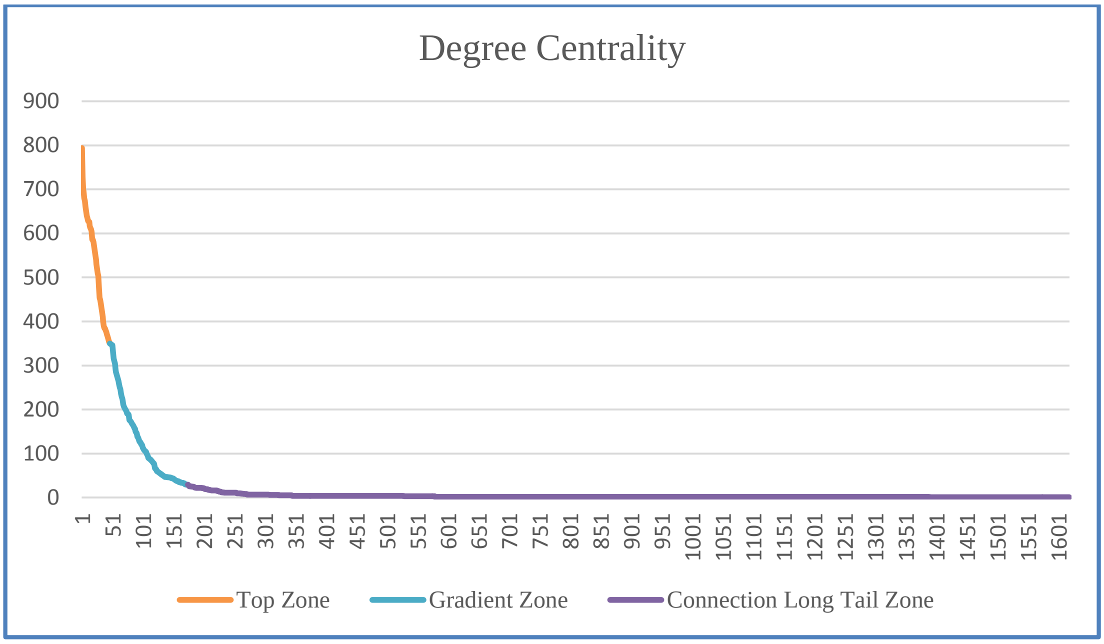


Figure 5-4: Degree Centrality of Diabetes Ontology (Tri-sectional trend of centrality distribution)

5.1.1.2 Definition Top Zone

Figure 5-5 demonstrates the degree centrality values for the top 47 concept terms in the diabetes ontology corpus. The statistical distribution of these predictions is close to a linear regression ($y = -8.8319x + 738.84$, $R^2 = 0.9804$). This regression indicates that the members of this zone share a rule concerning the variation of the degree centrality.

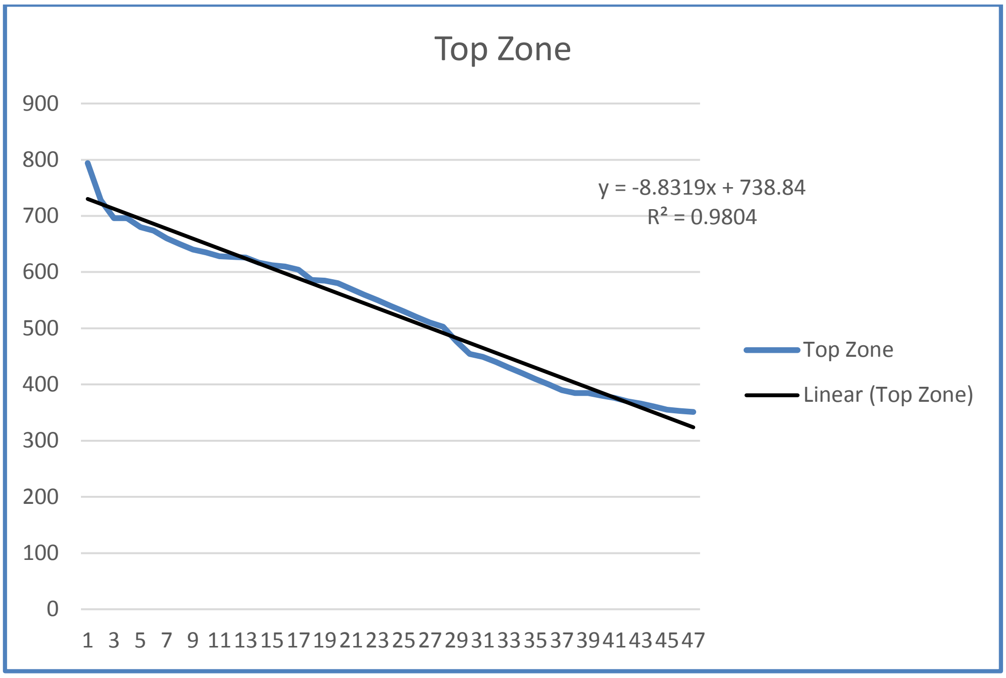


Figure 5-5: Centrality trend (Top Zone) in the diabetes ontology experiment

The derived Diabetes ontology top zone is populated by concepts with a high degree of centrality value. This indicates they are the most *commonly* nominated concept terms by other terms in the knowledge domain. Definition concepts within this zone are connected to each other directly. As a result, the remaining terms in the domain network that do not belong to this zone can still often find paths to other terms via the concepts in this zone. Hence, the concept terms in the top zone have the role of "root members" of the ontology and offer shared paths to all members in the Diabetes ontology.

Table 5-1: Top 20 Terms in the diabetes ontology

Top 20 Terms			
Prediction	**Centrality**	**Prediction**	**Centrality**
adult-onset diabetes	793	**Neonatal Diabetes Mellitus**	622
diabetes mellitus	727	**insulin-dependent diabetes mellitus**	622
gestational diabetes	695	**ketoacidosis-prone diabetes**	622
juvenile diabetes	695	**ketoacidosis-resistant diabetes mellitus**	618
type 1 diabetes	679	**ketosis-resistant diabetes**	618
type 2 diabetes	673	**ketosis-prone diabetes**	603
Hyperinsulinism	641	**ketosis-resistant diabetes mellitus**	585
adult-onset diabetes mellitus	630	**maturity-onset diabetes mellitus**	584
autoimmune diabetes	622	**non-insulin-dependent diabetes**	584
growth-onset diabetes	622	**Tolbutamide**	549

5.1.1.3 Gradient Change Zone

This gradient zone has 129 members with a degree centrality value ranges from 349 (latent diabetes) to 40 (Vegetarianism). The gradient (middle) zone members have a power type trend line ($y = 465.14x^{-0.206}$, $R^2 = 0.9689$), which almost matches their degree centrality distribution as shown in Figure 5-6. This match between the distribution and the trendline means this zone can be mathematically described.

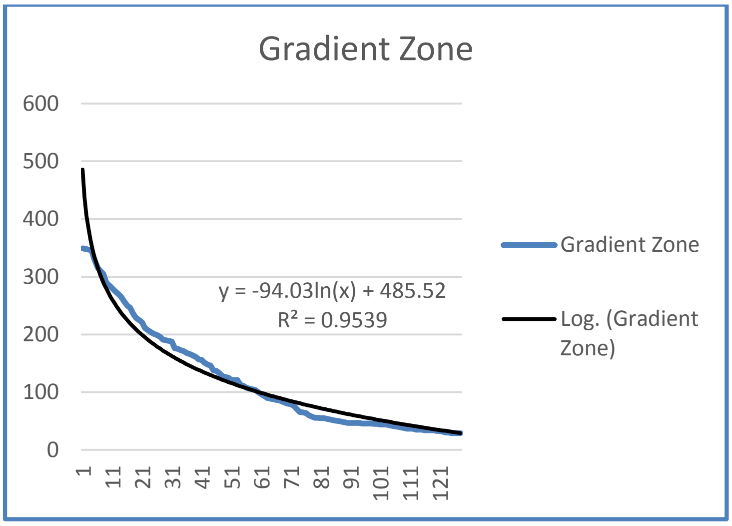


Figure 5-6: Centrality trend (Gradient Zone) in the diabetes ontology experiment

This zone contains "200" members, significantly more than "47" members contained in the top zone. This level contains popular concept terms that are inherited by the top-level concepts. Examining these concepts shows that members of this level are expressions or phrases that contain the top-zone terms, their synonyms, or relevant descriptors. Within this description level, the members are closely connected to the top-level concepts, yet they are not as highly central in the ontology network. For instance, "juvenile diabetes" is a top-level term in the derived Diabetes ontology. "Sugar diabetes" and "lente insulin" which are directly connected terms to "juvenile diabetes" are considered descriptors and hence exist in the middle description zone.

Each concept in the middle description zone has at least one direct relationship with top-level concept terms. The middle (description) level members have limited relations amongst themselves. This means that this level does not represent a fully connected sub-network within the ontology. That is a unique characteristic of top zone members.

The concepts in this zone can reach all concepts in the top definition zone and via them to any of the others in the description level within a maximum of three concept term traverses. The small number of traverses required shows that the middle level is strongly connected via key concepts in the top core zone. The members of this zone are more closely bound to the top zone compared with the ground "long tail" zone.

5.1.1.4 Long Tail Zone

The degree centrality trend of the remaining members in the diabetes ontology corpus has a power trendline regression ($y = 142.77X^{0.657}$, $R^2 = 0.9315$). Figure 5-7 shows the centrality distribution of these zone members.

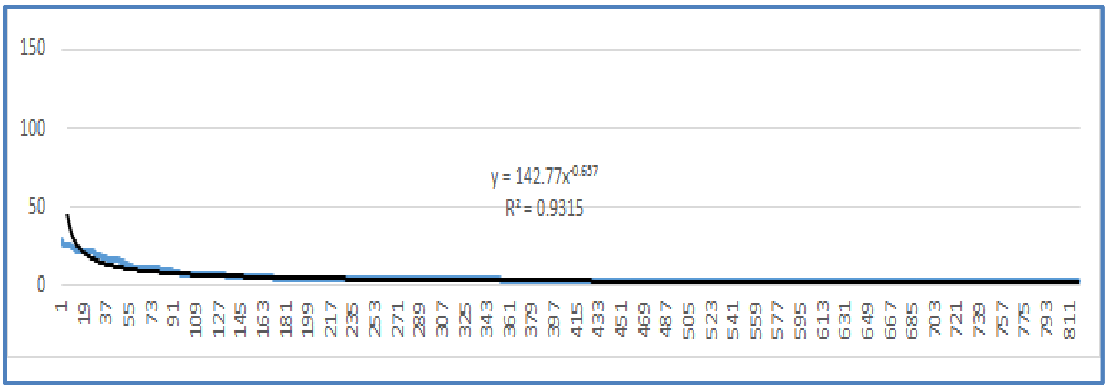


Figure 5-7: Centrality Trend (Connection Zone) in the diabetes ontology experiment

The number of members in this zone are significantly higher than the first two zones. The long tail zone members constitute around 75% of the total number of terms in the corpus. The average degree centrality of the members is 3.67, while the change rate of the degree centrality values in this long tail zone is "0.58%". This indicates that the long tail resembles an almost flat section of the distribution curve.

5.1.2 Closeness Centrality Analysis

Closeness Centrality was applied to the clusters of the diabetes ontology corpus found by the degree centrality analysis. For example, Table 5-2 shows the closeness centrality values $\mathbf{F_{cl}(t,S)}$ of the top zone concept terms to the keyword *diabetes mellitus*. The closeness centrality defines the relationship weight between two concept terms.

The column $\mathbf{F_{RD}(t,S)}$ is the relevant distance that shows how far the related concepts S are from t's point of view (in this case *diabetes mellitus*). This function takes the highest $\mathbf{F_{cl}(t,S)}$ as the representative of the closest member to the target term t. The bigger the value of $\mathbf{F_{RD}(t,S)}$, the further the corresponding member is from t in the cluster.

Table 5-2: Closeness Analysis – Diabetes Mellitus relationships

Seeding Keyword (S)	Target Term (t)	$F_{cen}(t,S)$	$F_{cen}(t)$	$F_{cl}(t,S)$	$F_{RD}(t,S)$
insulin shock	diabetes mellitus	71	689	0.1030479	1
gestational diabetes	diabetes mellitus	63	689	0.09143686	1.127
sugar diabetes	diabetes mellitus	38	689	0.05515239	1.8684
juvenile diabetes	diabetes mellitus	37	689	0.05370102	1.9189
type I diabetes	diabetes mellitus	37	689	0.05370102	1.9189
autoimmune diabetes	diabetes mellitus	37	689	0.05370102	1.9189
growth-onset diabetes	diabetes mellitus	37	689	0.05370102	1.9189
insulin-dependent diabetes mellitus	diabetes mellitus	37	689	0.05370102	1.9189
ketoacidosis-prone diabetes	diabetes mellitus	37	689	0.05370102	1.9189
ketoacidosis-resistant diabetes mellitus	diabetes mellitus	37	689	0.05370102	1.9189
ketosis-prone diabetes	diabetes mellitus	37	689	0.05370102	1.9189

Seeding Keyword (S)	Target Term (t)	$F_{cen}(t,S)$	$F_{cen}(t)$	$F_{cl}(t,S)$	$F_{RD}(t,S)$
ketosis-resistant diabetes mellitus	diabetes mellitus	37	689	0.05370102	1.9189
maturity-onset diabetes	diabetes mellitus	37	689	0.05370102	1.9189
non-insulin-dependent diabetes mellitus	diabetes mellitus	37	689	0.05370102	1.9189
Insulin	diabetes mellitus	37	689	0.05370102	1.9189
diabetes insipidus	diabetes mellitus	37	689	0.05370102	1.9189
Lente insulin	diabetes mellitus	37	689	0.05370102	1.9189
chemical diabetes	diabetes mellitus	37	689	0.05370102	1.9189
latent diabetes	diabetes mellitus	37	689	0.05370102	1.9189
nephrogenic diabetes insipidus	diabetes mellitus	37	689	0.05370102	1.9189
recombinant human insulin	diabetes mellitus	37	689	0.05370102	1.9189
adult-onset diabetes	diabetes mellitus	36	689	0.05224964	1.9722
type II diabetes	diabetes mellitus	36	689	0.05224964	1.9722
mature-onset diabetes	diabetes mellitus	36	689	0.05224964	1.9722
bronzed diabetes	diabetes mellitus	36	689	0.05224964	1.9722
insulin reaction	diabetes mellitus	36	689	0.05224964	1.9722
Polyuria	diabetes mellitus	34	689	0.04934688	2.0882
insulin shock therapy	diabetes mellitus	31	689	0.04499274	2.2903
Carbohydrate	diabetes mellitus	3	689	0.004354137	23.667
carbohydrate loading	diabetes mellitus	1	689	0.001451379	71

Several key attributes were identified by these analysis:

1. The closeness centrality analysis revealed that weighted relations among the members of the highly centralised top zone contributed to 9.87% of the total ontology relationships (635 out of 6431).

   In addition, 37.87% (2436 out of 6431) of the total relationships in the ontology is contributed by top zone keywords. Taking into consideration their high degree centrality, top zone members bring the other zones members closer to them.
2. The data in Table 5-2 shows that different keywords have different relationships to the same target term *diabetes mellitus*. It also proves that the relations are directed and weighted, which is a main advantage of this ontology.
3. Although the top zone members are not fully connected to each other, the cut off points' accuracy was not affected. Having the top zone members not fully connected might be a benefit to the analysis with regard to identifying concept clusters, this would affect the simplicity of the network and impose a significant Betweenness centrality calculation in the top zone because of the connection breaks in the zone. Therefore, as a result of the closeness centrality analysis, the members of the top zone that did not nominate any other member within the top zone were moved to the gradient zone.

5.1.3 Betweennesss Centrality Analysis

The members of the ontology with low $\mathbf{F_{cl}(t,S)}$ in the closeness centrality analysis were examined using the Betweenness centrality analysis. This analysis helps locate and identify members that might have been overlooked by previous centrality analysis, but can *bridge* different conceptual clusters. For example, Table 5-3 reveals that *type II diabetes* and *ketoacidosis prone diabetes* are not specifically close to each other.

However, there exists a member "gestational diabetes" that is closely connected to both concepts.

Table 5-3: Example of Betweenness Analysis in the Diabetes Ontology

Seeding Keyword (S)	Target Term(t)	$f_{Cen}(t)$	$f_{Cen}(t, S)$	$f_{CI}(t, S)$
type II diabetes	ketoacidosis prone diabetes	351	3	0.008547
ketoacidosis prone diabetes	type II diabetes	673	27	0.040118
gestational diabetes	type II diabetes	673	64	0.095097
gestational diabetes	ketoacidosis prone diabetes	351	50	0.142450

The ability to locate and identify members like "gestational diabetes" in the network has an important benefit from a practical perspective: In a "type II diabetes"-centred network, the devised ontology might not have included "ketoacidosis prone diabetes" in it. However, if the context of the application includes "gestational diabetes", then "ketoacidosis prone diabetes" should be considered as an option to be included.

5.1.4 Diabetes Ontology Result Summary

The three zones of the derived ontology have the ability to model structurally the target domain knowledgebase: Core, related and peripheral concepts have been identified by the process derived by this research. However, the content of these zones needs to be evaluated.

5.1.5 Ontology Evaluation

Integrating various evaluation methods proposed in the literature, such as METHONTOLOGY evaluation method, which proposes an ontology evaluation phase from a knowledge representation perspective (Gómez-Pérez, 2001). This method evaluated ontology processes and the output concepts in order to assess the internal structure and relations compared to actual practical knowledge structure.

Content validation (Welty and Guarino, 2001) and practical evaluation (Staab et al., 2001) are also used in this research to evaluate the ontology. The former evaluation has detailed methods for the assessment of an ontology engineering approach, according to its output, while the latter emphasises that the evaluation should consider the opinions and requirements of the ontology users. Each one of these evaluations could aid the usability justification of the derived ontology. However, they may overlook or ignore concerns that may arise through other evaluation approaches. Hence, it is proposed to assess the derived ontology from multiple aspects (Figure 5-8). Content Evaluation (Completeness and Conciseness) and a practical evaluation were conducted.

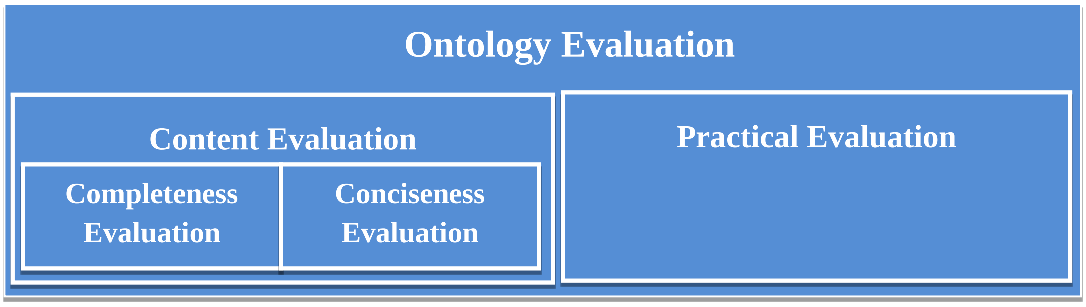


Figure 5-8: Ontology Evaluation approaches

5.1.5.1 Diabetes Ontology Content Evaluation

This evaluation examines the validation of the concepts and relations of the derived ontology. This section examines two main aspects of content evaluation, namely the Ontology Content Completeness and the Ontology Content Conciseness of the derived ontology network.

5.1.5.1.1 Ontology Content Completeness Evaluation

This evaluation concentrates on examining the coverage of the ontology and whether it has generated a sufficient number of terms in order to represent the target domain. It reflects how *adequate* the ontology coverage of the domain is.

However, it is really impossible to show if any domain ontology is complete because of the emergence of new knowledge in nearly all domains. Hence, checking the content completeness of the ontology is usually conducted as an examination of the content *in*completeness. This means that the derived ontology is scanned to discover the *missing* terms. Missing terms are the ones that should have been incorporated in the ontology corpus or the ones that can be deduced from other terms.

The derived ontology was compared with the diabetes ontology (OGMD) and the “Disorder of glucose metabolism” section of SNOMED CT in order to conduct a completeness evaluation. However, since the long tail zone terms are peripheral to the core ontology, the analysis will only examine the core members of the ontology.

Table 5-4 shows the comparison between the top level of the derived diabetes ontology produced in this research, with the “OGMD” ontology and “SNOMED CT” ontology. The first column of the table shows the concepts from “OGMD” ontology; the second column of the table shows the “SNOMED CT” ontology, while the last two contain the concepts and their degree centrality for the derived diabetes ontology. The degree centrality reflects the level to which the concept belongs. The different levels were marked using different background colours; Green-coloured rows show terms from the top level and Blue-coloured rows show terms from the middle gradient level.

Table 5-4: Top Definition Level Concepts in the Diabetes Case study

OGMD	SNOMED	New Diabetes Ontology	Matches
Diabetes Mellitus	Diabetes Mellitus	Diabetes Mellitus	727
Hyperglycemia	Hyperglycemia	Hyperglycemia	258
Hyperinsulinism	Hyperinsulinism	Hyperinsulinism	641
Hypoglycemia	Hypoglycemic Syndrome	Hypoglycaemia	86
Prediabetes syndrome		Prediabetes	136
Diabetes Mellitus syndrome in newborn infant	Neonatal diabetes mellitus	Neonatal Diabetes Mellitus	622
	diabetes mellitus maturity onset	maturity-onset diabetes mellitus	
Diabetes Mellitus, Insulin Dependent	insulin-dependent diabetes mellitus	insulin-dependent diabetes mellitus	622
Diabetes Mellitus, Non-Insulin-Dependent		non-insulin-dependent diabetes	584
Gestational diabetes complication pregnancy, childbirth, or the puerperium	Gestational diabetes mellitus	gestational diabetes	695

Top Definition Level Middle Description Level

The OGMD ontology has proposed the use of seven concepts in its top level (First seven rows in Table 5-4): *“Diabetes complication”*, *“Diabetes Mellitus”*, *“Hyperglycemia”*, *“Hyperinsulinism”*, *“Hypoglycemia”*, *“Metabolic Disorders associated with hypertension and diabetes”*, and *“Prediabetes syndrome”*. The generated diabetes ontology includes five out of seven of those terms in its top level, while two of them are included in the middle level with high degree centrality values. This might be due to the fact that OGMD has been developed by domain experts and reflects their understanding of the domain and the application requirement, while the automatic generation of the derived diabetes ontology has relied on the general understanding of the domain reflected in the knowledgebase. It is argued that two groups of experts could also have achieved this level of differentiation.

The new derived ontology contains more concepts compared to both OGMD and SNOMED. Table 5-5 shows extra top-level members that belong to the new ontology, which were not included in the "OGMD" diabetes ontology. This broad coverage of the new derived ontology shows that the new diabetes ontology is able to represent "OGMD" terms, yet still be able to provide more breadth and depth of coverage.

Table 5-5: Extra Concepts in the Top-level zone of the new Diabetes Ontology

Extra Concepts (with their centralities)					
adult-onset diabetes	793	ketoacidosis-resistant diabetes	618	Acetonuria	502
juvenile diabetes	695	ketoacidosis-resistant diabetes mellitus	618	banting	476
type I diabetes	679	ketosis-prone diabetes	603	acetonemia	453
type II diabetes	673	ketosis-resistant diabetes mellitus	585	ketoaciduria	448
adult-onset diabetes mellitus	630	mature-onset diabetes	584	ketonemia	436
autoimmune diabetes	622	maturity-onset diabetes mellitus	584	insulin	419
growth-onset diabetes	622	Tolbutamide	549		
ketoacidosis-prone diabetes	622	Diabetic	547		

In addition to these concepts, OGMD and SNOMED CT have presented further concepts and descriptions for diabetes mellitus. The new derived ontology did not include all of these concepts and descriptions. Overall, more than 75% of the OGMD ontology and the SNOMED CT's "Disorder of glucose metabolism" section have appeared in the new ontology derived. This can be justified as OGMD was specially developed for a specific application with a higher correspondence to the genetic susceptibility to diabetes. As the new derived ontology was trying to get the concepts and relationships using three starting keywords, specific terms and descriptions related to OGMD specific application were weighted less by the algorithm developed.

The coverage of the ontology derived includes additional concepts that might be considered as redundant, which are not as specifically connected as the core concepts. This could challenge the accuracy of the defined terms and their connections. This creates the requirement for an Ontology Content Conciseness Evaluation.

5.1.5.1.2 Ontology Content Conciseness Evaluation

Content conciseness evaluation is different from the completeness evaluation as it concentrates on cutting the additional terms that can be deemed redundant in order to shape the ontology with reference to the minimum practical set of terms and their descriptors.

Traditional ontology and terminology such as OpenGALEN (discussed in 2.3.1.2) attempt to restrict redundant terms and connections in order to form the minimum viable group of concepts and relationships. However, this intolerance of redundancy of concepts and relations restricts the ontology's coverage of the subject area and allows concept gaps to appear between the terms. Therefore, strict application of the content conciseness could weaken the derived ontology with regard to the goals of this research. It is understandable that the content of a human-oriented ontology can be concise, since the human reasoning can be utilised to bridge the semantic gaps in the ontology, whereas in an ICT System-oriented ontology, the gaps are harder to be traversed since there is a lack of logical reasoning available compared to human experts. Therefore, it is suggested that some level of concept redundancy in order to meet the conditions of broad coverage of the domain area and content conciseness be tolerated.

The use of weighted and directional connections can help in addressing the problem of measuring the concept of redundancy. If the ICT system requires a specific level of conciseness, a pruning process can be implemented to eliminate those concepts that have a relation value with the top definition level that is smaller than the conciseness value required. For instance, in the new diabetes ontology, if there was no specific conciseness requirement, all the concepts in the top definition level have a fully connected sub-network. Hence, any term from this network should direct to all the nodes in the sub-network. However, if a conciseness requirement was to set the relationship weight limit to $f_{RD} \leq 1.86$, then a “type II diabetes”-centred sub-network would only contain the concepts that can satisfy this requirement (Figure 5-9).

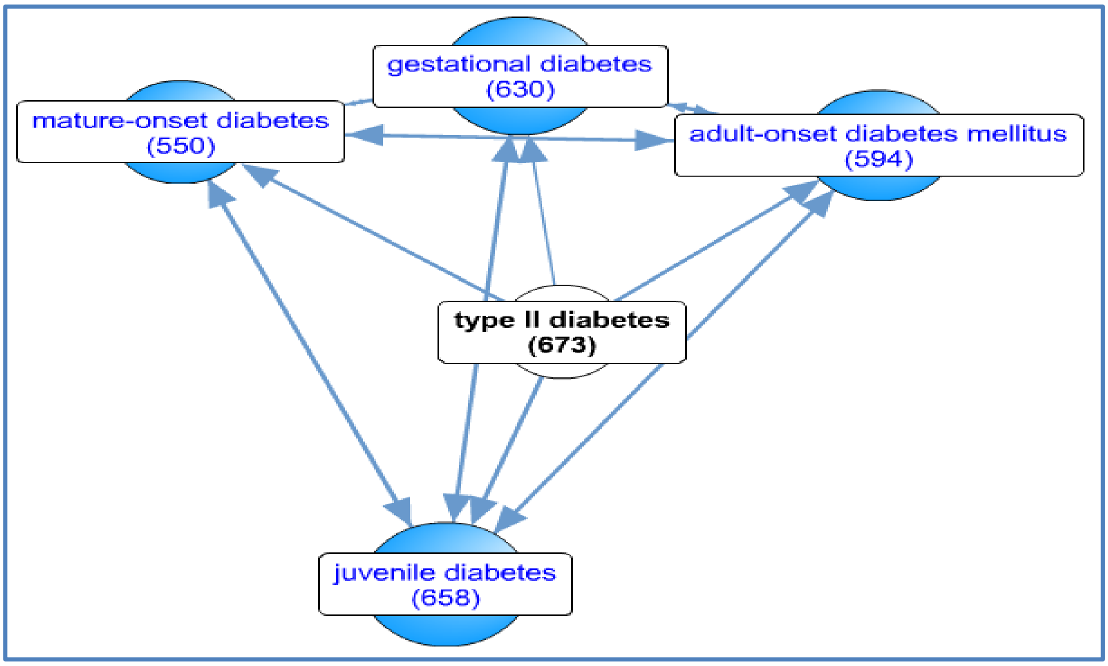


Figure 5-9: "type II diabetes"-Centred Network

Even though the new derived diabetes ontology is not concise in the broad sense, the ability to configure the resultant ontology provides the flexibility of using the same result in different ICT systems that have different conciseness needs.

Having evaluated the new derived ontology for both content completeness and content conciseness, a practical evaluation was carried out by integrating the domain ontology within the proposed text mining method.

5.2 Semantic Text Mining Results

In order to analyse the Diabetes.co.uk forum text, the content of the forum was downloaded in HTML format from its website. This was carried out using website downloading software, WinHTTrack (www.httrack.com), in order to capture the Diabetes Forum discussions for analysis. This program is open source software, which allows the user to download the specified forum to a local folder for further analysis. Figure 5-10 shows the WinHTTrack interface.

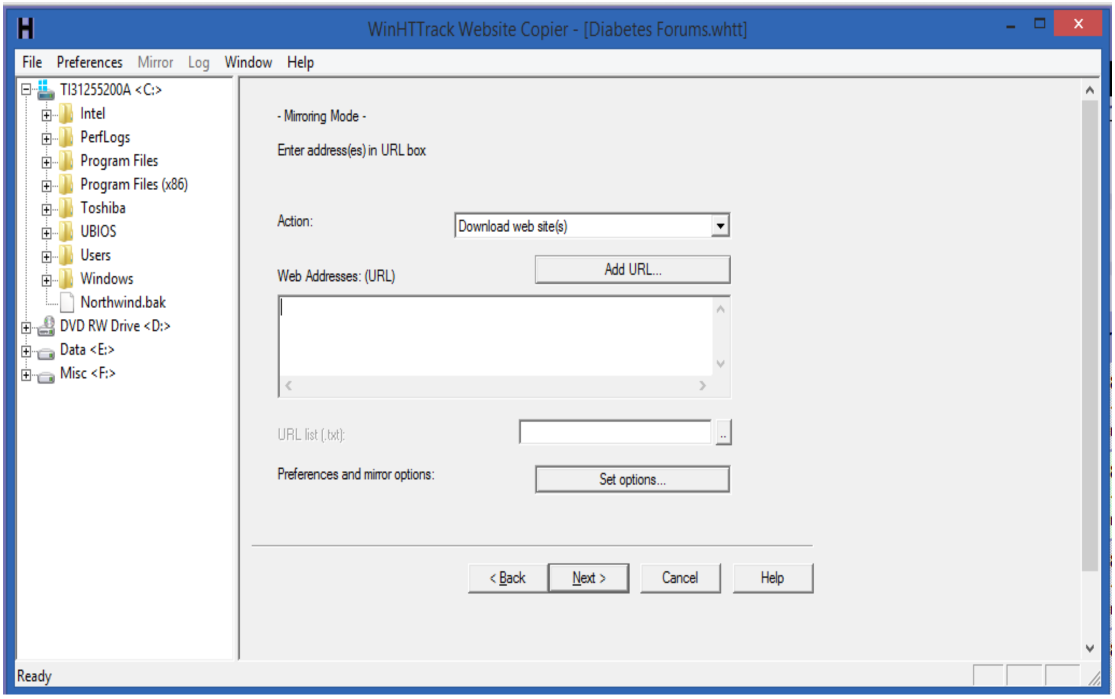


Figure 5-10: WinHTTrack Interface

Forum's threads and posts are then extracted from the downloaded HTML files, and they are stored in a local repository for further text processing.

The data for this research was downloaded at the start of 2012 and updated to include all the threads up to 01/04/2013 from the diabetes.co.uk forum[13]. The diabetes forum had 239,668 posts between January 2007 and end of March 2013. These posts were organised in 25,388 threads.

Table 5-6 shows the descriptive statistics of the data obtained.

[13] http://www.diabetes.co.uk/forum/

Table 5-6: Description Statistics of Diabetes Forums

Description	Result
No. of Threads	25,388
No. of Posts	239,668
Minimum no. of posts in a thread	1
Maximum no. of posts in a thread	1326
Average no. of posts in a thread	9.44

The number of posts has grown over the first six years from 26855 in 2008, to 78755 posts in 2012 (Figure 5-11). This growth may reflect the increasing interest in using social media by patients to search for and discuss information related to their condition online.

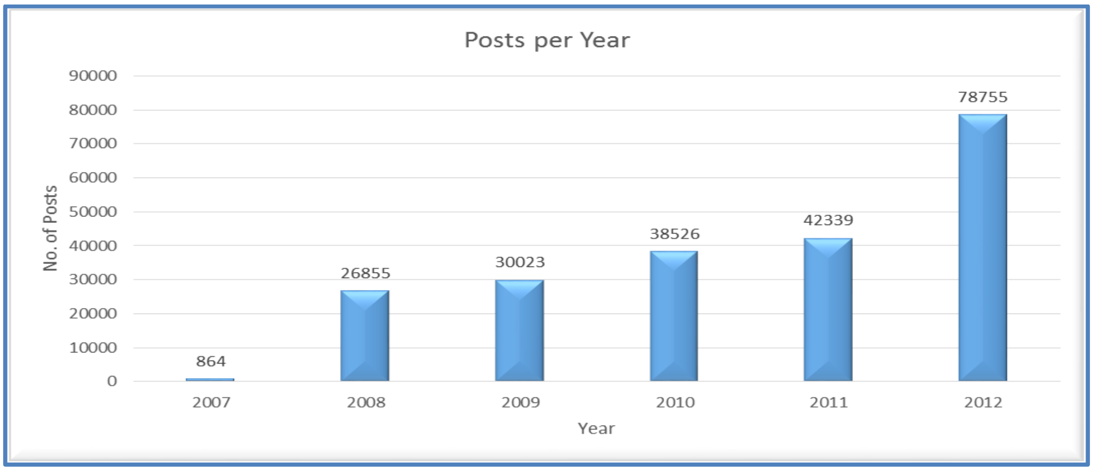


Figure 5-11: Number of posts per year

The extracted threads and their posts were passed to the first component in the text mining method – Semantic Annotation and Topic Identification in order to enrich and annotate the text, as well as identify the topics in the threads.

5.2.1 Semantic Annotation and Topic Identification Results

The process described in Section 4.1.1 takes as an input the diabetes forum text as an unstructured textual dataset that consists of hundreds of thousands of posts, then semantically annotates it with the diabetes ontology and DBpedia ontology using the terms found in these posts.

A semantic annotation software component has been designed and implemented in order to implement the proposed model of combining both the domain ontology with DBpedia Spotlight in order to annotate the diabetes forums' posts. It should be noted that in this research, the posts are the primary analysis element, i.e. the post is treated as the basic unit of analysis for identifying entities and their semantic relations.

From these posts, more than 3000 terms and entities were extracted and semantically annotated by the derived diabetes ontology and DBpedia Spotlight. Many terms were commonly found by both of these sources of knowledge, such as "Diabetes Mellitus" and "Prediabetes". However, other terms such as "aleuronat flour" and "hyperinsulinism" were found by the diabetes ontology only, while other terms such as "Polycystic ovary syndrome (PCOS)" were annotated by DBpedia alone. This highlights the importance of the combined hybrid annotation system and its ability to capture entities and annotate the text in a more effective way than using either sources on their own in order to enrich the text for further processing.

Table 5-7 shows descriptive statistics of the annotations obtained. On average, the annotation frequency was 197.42. The highest frequency of an annotation was "Blood Sugar" with 39,894 posts annotated with it. Whereas 245 terms shared the minimum frequency of 10, examples include "Aromatherapy", "Béchamel sauce" and "Compact Fluorescent Lamp". Even though these concepts had low frequency, they were of concern to the forum users. For example, Compact Fluorescent Lamps (CFLs) were discussed in the forum after a news agency in Canada has reported a study that links CFLs and diabetes. The study by Havas (2008), in fact, was reporting that transient electromagnetic field (TEF) may be "contributing to elevated blood sugar levels among diabetics and pre-diabetics". The discussion around CFLs decreased when this information about the study was shared by one of the users.

Table 5-7: Annotation Descriptive Statistics

Statistical Measure	Value
Average Annotation Frequency	197.42
Maximum Annotation Frequency	39,894
Minimum Annotation Frequency	10
Standard Deviation of Annotation Frequency	1284.1

The annotations resulted from enriching the text using the external sources reflect a wide range of concepts. For example, these concepts might be a diabetes drug such as *Glucophage (Metformin)*, a food type such as *Fruit* or sometimes they can be a term that describe the connection between two other concepts, such as a *complication* or *side effect*.

The resulting annotations from the semantic annotation stage were used to identify the topics of the threads within the forum. For each thread, the associated annotations of the posts were retrieved. Then, a SPARQL query was used to navigate DBpedia and obtain the "subject" of the annotation as well as broader relation between the annotation and its topics. For example, an annotation of "Low carbohydrate diet" is under the category "Diets", which in turn under the category "Nutrition".

Table 5-8: Top 10 Frequent Annotations

Annotation	Frequency
Blood sugar	39894
Carbohydrate	34110
Diabetes mellitus	31964
Food	18834
Insulin	18791
Diabetes mellitus type 2	14577
Diet (nutrition)	13301
Eating	12199
Meal	11721
Glucose	8070

It was observed that many terms and their topics belong to several high-level categories, such as “diabetes drugs” and “Foods and Nutrition”. Each of these categories contains a significant number of entities, which indicates their importance to the forum members.

Table 5-9 summarises the observed categories and their size. What is also important about these analysis-based categories is that they correlate to the existing discussion themes within the diabetes community. In this way, these generated categories provide a structure for the entities and topics extracted from the posts. This facilitate the analysis of these categories and the relationships between the entities within them.

Table 5-9: The size of generated categories from the analysis of diabetes forums

Category	No. of Entities	No. of Posts	Examples
Body organs	47	5455	Pancreas, Heart, Gallbladder
Complications	71	7820	Ketoacidosis, Glaucoma, Nephropathy
Diabetes Types	13	4046	Type I and II, LADA, Diabetes Insipidus
Diabetes Drugs	36	23076	Metformin, Lantus, NovoRapid, Levemir
Diabetes Diagnosis and Symptoms	95	30129	Polydipsia (thirst), c-Peptide, Candidiasis
Food and Nutrition	426	79532	Carbohydrates, LCHF Diet
Social Aspects and Lifestyle	61	24083	gym, exercise, fitness, music, yoga

The category *Food and Nutrition* is the biggest category with 30% of entities. Having the *Food and Nutrition* at the top reflects that nutrition related discussions are very common and constitutes a significant issue for diabetics. *Diagnosis and Symptoms* is next with 12%. *Social Aspects and Lifestyle* are in the third place with 10%.e. These three categories echo the main concerns amongst the forum members. Analysing these categories and the relationships between them could provide interesting insights into the different concerns of their daily life with diabetes.

5.2.1.1 Semantic Annotation Evaluation

In order to practically evaluate the semantic annotation component, 200 random posts were manually annotated by the author with the help of a diabetes expert from Warwick Medical School. This was necessary because there is no standard forum dataset to enable the evaluation and comparison between different annotation systems.

The precision and recall measures were calculated using these manually annotated posts to test the performance of this component. Recall represents the percentage of correctly-annotated entities out of the overall entities in the text. Precision represents the percentage of correctly-annotated entities out of the retrieved entities. The formulas to calculate both measures are as follows:

$$\text{Precision} = \frac{\{\text{Relevant Annotations}\} \cap \{\text{Retrieved Annotations}\}}{\text{Retrieved Annotations}}$$

$$\text{Recall} = \frac{\{\text{Relevant Annotations}\} \cap \{\text{Retrieved Annotations}\}}{\text{Relevant Annotations}}$$

For example, a manually-annotated dataset contains 90 entities in total; The annotation process identified 70 entities; Out of these 70, only 55 entities were correctly annotated. Then the Recall in this example is: $\frac{55}{90} * 100 = 61.11\%$ and the precision is $\frac{55}{70} * 100 = 78.57\%$. This indicates that the system favours precision over recall and that it is more concerned with the quality of the annotations than their quantity.

A commonly used measure to combine both recall and precision is the F-score, which is calculated as follows: $F = 2 * \frac{precison * recall}{precision + recall}$

F-score measures the performance accuracy of a text mining system, and represents the ***sub-contrary mean*** (Sokolova et al., 2006) of precision and recall. Sub-contrary mean is a type of average measures that is used for calculating the average of two rates. In the previous example, F-score can be calculated as follows:

$F = 2 * \frac{\frac{61.11}{100} * \frac{78.57}{100}}{\frac{61.11}{100} + \frac{78.57}{100}} = 68.74\%$. This F-score value indicates a relatively accurate system.

The analysis of the manually-tagged posts has resulted in a Recall value of 78% and precision of 91%, which leads to F1= 84%. This result was compared to the annotation results from the derived domain ontology, as well as comparing the results to two other studies (Netzer et al. (2012) and Majumder and Saha (2014)) that used machine learning and training data to extract terms from web content, such as customer reviews. Table 5-10 shows the results and comparison. The performance of the chosen approach was better than the other systems, especially in recall measure. This probably reflects the value of using a hybrid annotation system for analysing the text in the forum. However, the comparison with other studies may not truly reflect the actual performance of the system as the datasets are different. This is also true for the results of other two studies in the table as the evaluation measures were derived through comparing the system performance against manually annotated posts.

Table 5-10: Evaluation of Semantic Annotation

	Domain Ontology	Domain Ontology + DBpedia	Netzer et al. (2012)	(Majumder and Saha, 2014)
Recall	70%	78%	74.4%	77.29%

Precision	80%	91%	90.3%	90.61%
F1	74.67%	84%	81.6%	83.42%

The output of the semantic annotation phase was then used to carry out further analysis using co-occurrence networks and dynamic tag clouds.

5.2.2 Dynamic Tag Cloud Results

The output of the semantic annotation and topic identification process discussed in the previous section was used as an input to the dynamic tag cloud generator. The annotations and their associated text extracted from the posts (known as Surface Form) were grouped based on the posts' dates on a monthly basis. Each group represents a month and is used to create the tag cloud for that period.

Figure 5-12 shows a calculated importance curve for demonstrating the overall view of the forums' content development over five-year time from 2008 to 2012. From the importance chart, it is observed that the high importance value starts from March 2008 compared with importance values before it. This was due to the appearance of new discussed topics, especially food and diets, such as the low-carb diet. After that, the topics about treatment, food and symptoms stayed as the dominant topics. However, several importance peaks can be observed in "May 2009" and "May 2010" (see Figure 5-13). In the former date, the discussion included new topics around Low-carb diet, thrombosis and Warfarin (an anticoagulant normally used in the prevention of thrombosis and thromboembolism), while in the latter, discussions around "sleep apnoea", foods such as "Black pudding" and alternative therapies such as "music" were introduced. (b) May 2010

Figure 5-13 (a)-(b) shows the corresponding tag clouds.

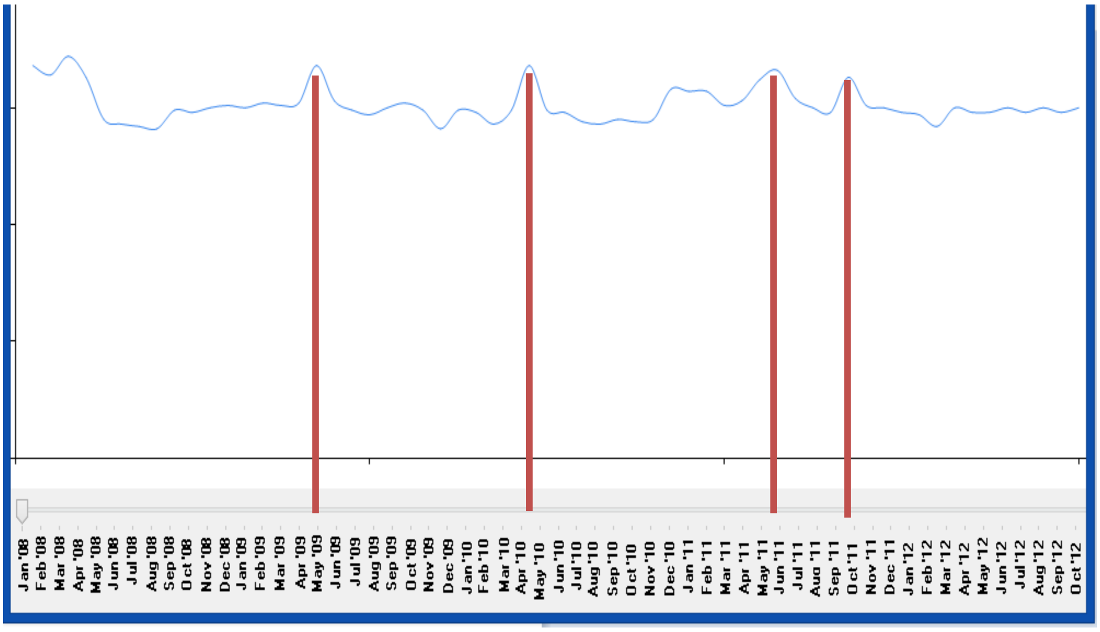
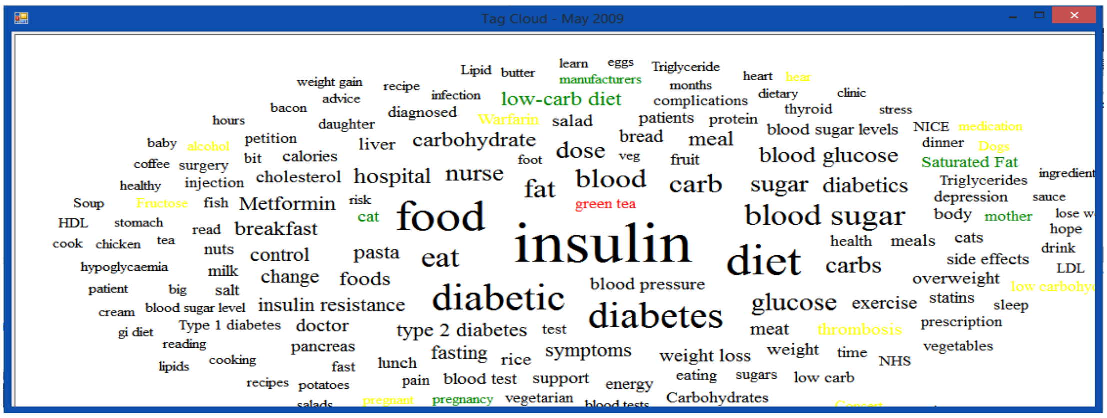
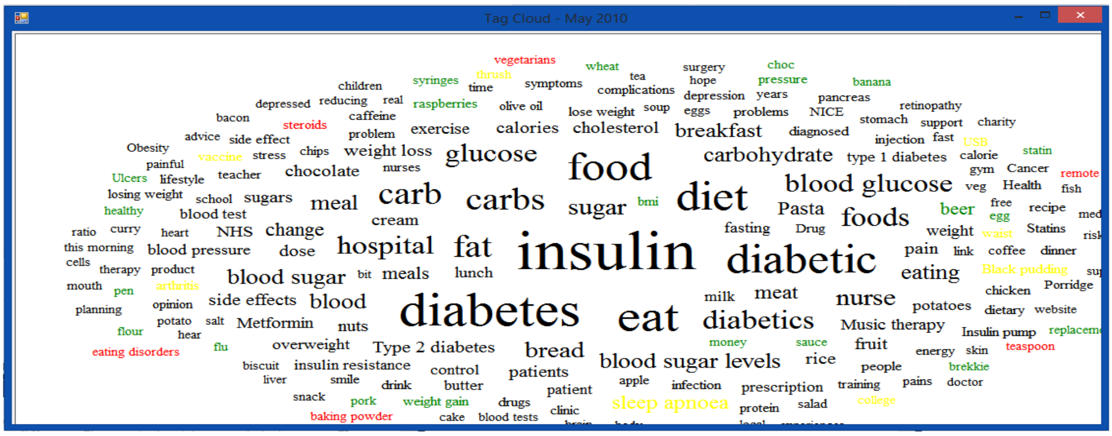


Figure 5-12: Dynamic Tag Cloud – Importance Curve



(a) May 2009



(b) May 2010

Figure 5-13: Dynamic Tag Cloud

There are also other interesting patterns, which can be found in these tag clouds. For example, when tracking the frequency (font size) of “LCHF diet” (Figure 5-14), its pattern shows two main peaks around June 2012 and October 2012. When comparing this chart with Google trends for “LCHF” diet, it turns out that this diet became popular in UK from 2011, with particular interest in it during 2012 and 2013 in the news media.

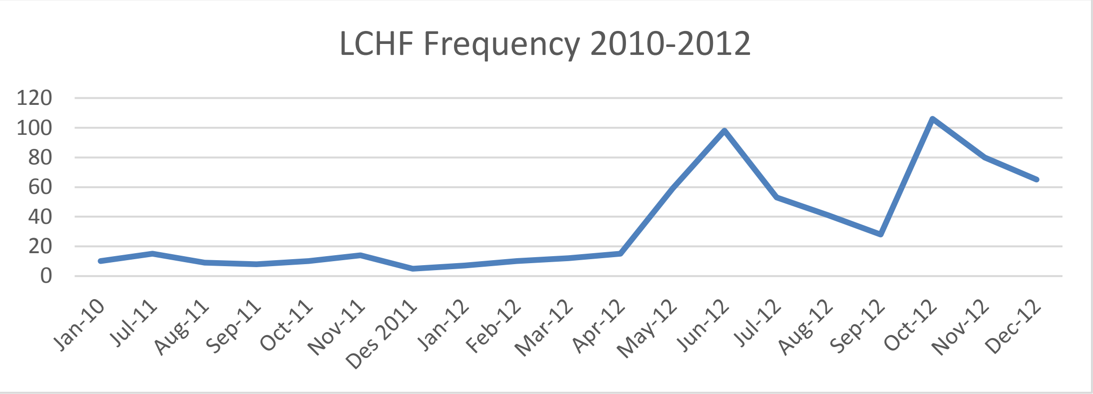


Figure 5-14: Frequency of “LCHF” in diabetes.co.uk forums between January 2010 and December 2012

The results of the dynamic tag cloud and importance curve have demonstrated that this layout could provide a novel method of summarising the text visually and highlighting interesting patterns or topics within the forum corpus. Yet, the relationships between terms identified in the dynamic tag cloud were not clear enough. The use of co-occurrence network could help identify interesting trends within the forum as a whole, or around a particular term.

5.2.3 Diabetes Forums Co-occurrence Results

In order to build the co-occurrence network, a Pointwise Mutual Information (PMI) measure is calculated between the terms and entities extracted from the forum. The more frequently two terms co-appear in the same post, the closer these entities are in the co-occurrence network. As discussed in section 4.3, the concept behind this linkage comes from the associative network concept. Whether the association between the entities and terms is positive, neutral or negative, the link between the terms is valuable to guide investigation. It is assumed that since the patient has used the two entities together, there is a relationship or closeness between these two entities in their mind.

A matrix of PMI values between annotations observed in the discussion forum was constructed according to Equation 4-7. The relationship between two nodes in the resulting network is undirected (symmetric) and the weight of the relationship is the PMI value. Examining the features of this co-occurrence network could help reveal interesting patterns of discussion within the forum.

As the network contains more than 3000 terms and concepts, it would be impractical to visualise and obtain meaningful patterns from such a large network. Therefore, sub-networks can be identified and analysed based on the categories obtained in the topic identification component. Hence, a parameterised co-occurrence network can be built and examined based on those categories. For example, in the diabetes forum case, a sub-network can be built to investigate the relationship between terms within "Food and Nutrition" category. Then, the relationships between terms in different categories can be examined. In the following analysis, two co-occurrence network for "Diabetes Drugs" and "Food and Nutrition" categories are analysed, then a concept-centred network is generated around a "Low Carbohydrates, High Fat" (LCHF) diet. This concept-centred network has the potential to reveal interesting associations with that concept.

5.2.3.1 "Diabetes Drugs" Category

As the forum contain the patients' experiences with using Diabetes Drugs, it is useful to investigate the discussions that contain diabetes drugs brands (the *Surface Forms* of the drugs) with other terms in the network. This would provide insights into different aspects of the diabetes management and gain an understanding of their views on the diabetes drugs.

Thirty five drugs brands were identified using the text mining method. Table 5-11 lists the extracted drugs, their frequency in the posts and the associated annotation. *Lantus* was the highest discussed drug name in the forum in 3709 posts. However, *Metformin* as a generic drug name was the highest discussed generic drug with 9275 mentions in the posts. This is expected since *Metformin* is the first drug choice for type II diabetes.

Table 5-11: Extracted diabetes drugs from diabetes.co.uk forums

Surface Form	Frequency	Annotation	Surface Form	Frequency	Annotation
Lantus	3709	Insulin glargine	**Amaryl**	37	Glimepiride
NovoRapid	2261	Insulin aspart	**Starlix**	32	Nateglinide
Levemir	2020	Insulin detemir	**Glipizide**	30	Glipizide
Byetta	1811	Exenatide	**Insuman Basal**	26	NPH insulin
Gliclazide	1546	Gliclazide	**Lucentis**	21	Ranibizumab
Victoza	960	Liraglutide	**GlucaGen**	21	Glucagon
Humalog	825	Insulin lispro	**Janumet**	11	Sitagliptin
Glucophage	447	Metformin	**Glucobay**	11	Acarbose
Apidra	407	Insulin glulisine	**Symlin**	10	Pramlintide
Januvia	276	Sitagliptin	**Galvus**	9	Vildagliptin
Actrapid	139	Insulin	**Eucreas**	8	Vildagliptin
Bydureon	121	Exenatide	**Prandin**	7	Repaglinide

Surface Form	Frequency	Annotation	Surface Form	Frequency	Annotation
Glimepiride	114	Glimepiride	**Competact**	6	Pioglitazone
Actos	110	Pioglitazone	**Tolbutamide**	5	Tolbutamide
Humalog Mix	75	Insulin lispro	**Onglyza**	3	Saxagliptin
Humulin S	70	Humulin	**Minodiab**	2	Glipizide
Humulin I	68	Humulin	**Regranex**	1	Becaplermin
Diamicron	49	Gliclazide			

Figure 5-15 shows the co-occurrence network for diabetes drugs.

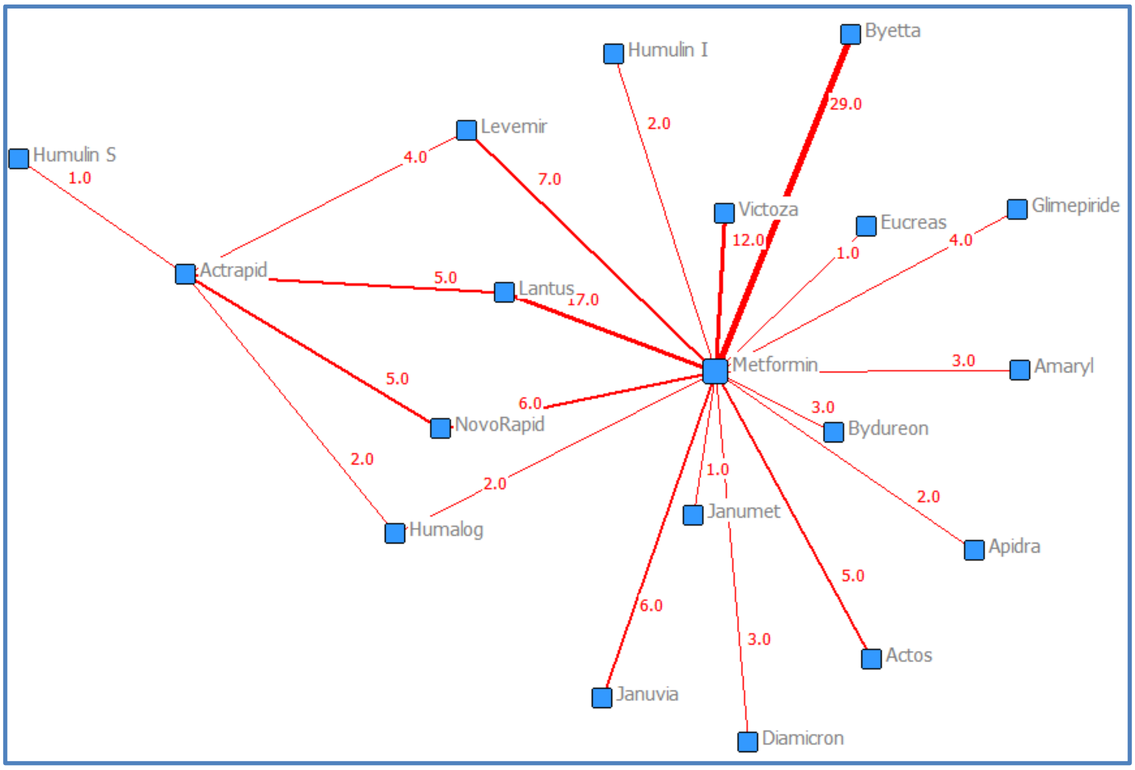


Figure 5-15: Diabetes Drugs Co-occurrence Network

In this network, each node represents a keyword using a name with a number: The number indicates how *popular* the keyword within the network, i.e., the number represents the appearance of the keywords in this network (Centrality). The edge thickness represents the absolute weight of the link between the two nodes. The edge length represents the relevant distance of the nodes based on the selected node. Viewing this network from metformin's perspective results in Figure 5-16 network:

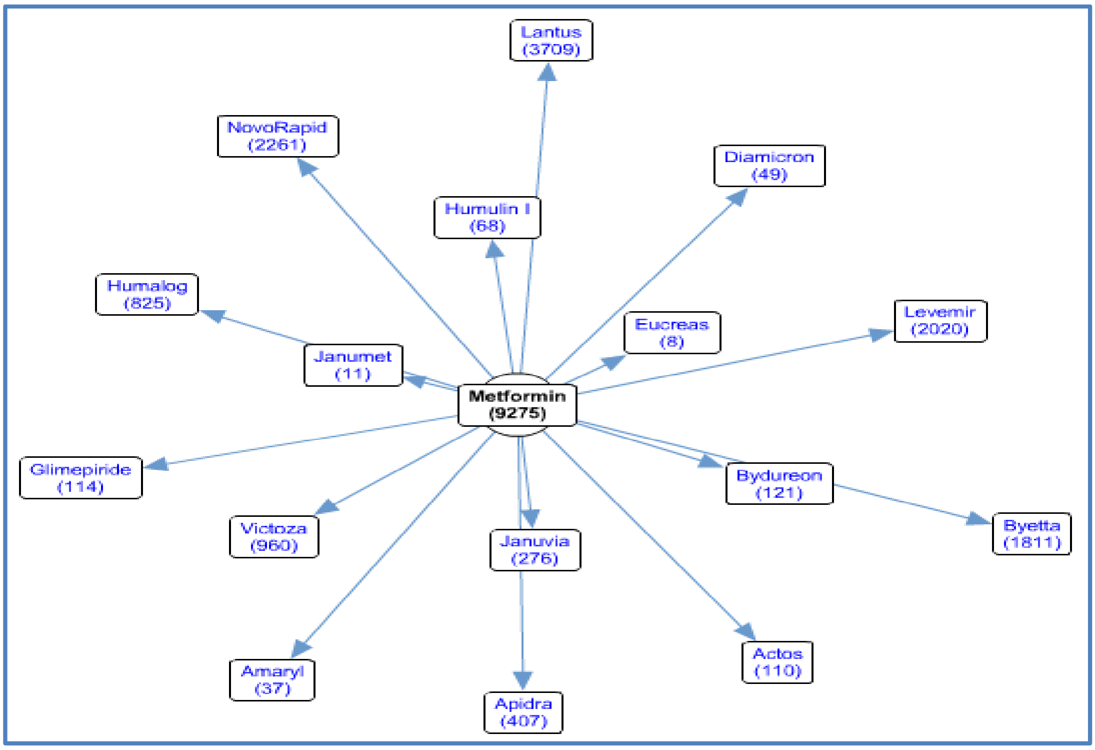


Figure 5-16: Metformin centred network

It can be observed from Figure 5-15 and Figure 5-16 that metformin is the central drug in this network, since all other drugs have co-appeared with it in the forum. It shows its dominance in the discussions around diabetes drugs. Another interesting observation is that *Actrapid* was the second central drug when comparing drugs in the same post, even though it had lower frequency (139) compared with other drugs. Since the diabetes drugs co-occurrence was low, further investigation was carried out to identify the reason behind it. It turned out that diabetes forum members prefer to compare generic types of drugs instead of the brand names. For example, they might compare rapid analogue insulin (e.g. Novo Rapid) with regular human insulin (e.g. Actrapid) instead of comparing the brand names.

5.2.3.2 Co-occurrence Network for "Food and Nutrition" Category

The Food and Nutrition category contains 426 terms that were grouped in subcategories using their corresponding topics, such as *Meat*, *fruit* and *vegetables*. Table 5-12 shows the Food and Nutrition category and its subcategories.

Table 5-12: Food and Nutrition Category

Category	No. of unique Annotations	Examples
Fruits	39	Banana, Apples, Mango, Pineapple
Vegetables	64	Broccoli, Cabbage, Chili Pepper, Legume, Rhubarb
Meat and Seafood	59	Beef, Bacon, Mackerel, Salmon
Nutrition	42	Cholesterol, Carbohydrates, Glycemic index, Mediterranean diet
Misc.	220	Burger King, Pizza, chocolate
Total	426	

Even though Food and Nutrition category contained a large number of terms, the resulted co-occurrence matrix was sparse (16.3%). This means, on average, 16% of Food and Nutrition term pairs have appeared in the same post at least once. The most *co-occurred*-with term in this category posts was "Bread" as it has "357" links with other terms. Social Network Analysis has been used to calculate the degree centrality of different Food and Nutrition terms to the discussion. “Carbohydrates” and “Glucose” were the top mentioned entities in this category network, but the “Bread” and “Carbohydrates” were the highest in terms of degree centrality. Other annotations that had lower frequency, such as “Bacon” (30th), “Biscuits” (31st) but had a high degree centrality measure were 4th and 8th respectively in terms of degree centrality. These are foods that tend to be discussed and compared against other types of food by diabetics and, hence, they have a higher degree centrality in the network.

Figure 5-17 shows a visual representation of the co-occurrence network for the Food and Nutrition category using the PMI method for calculating the relationships between the terms in those categories. The size of the nodes in this graph represents the frequency of the term in the forum, while the length between the relationships represents the PMI value.

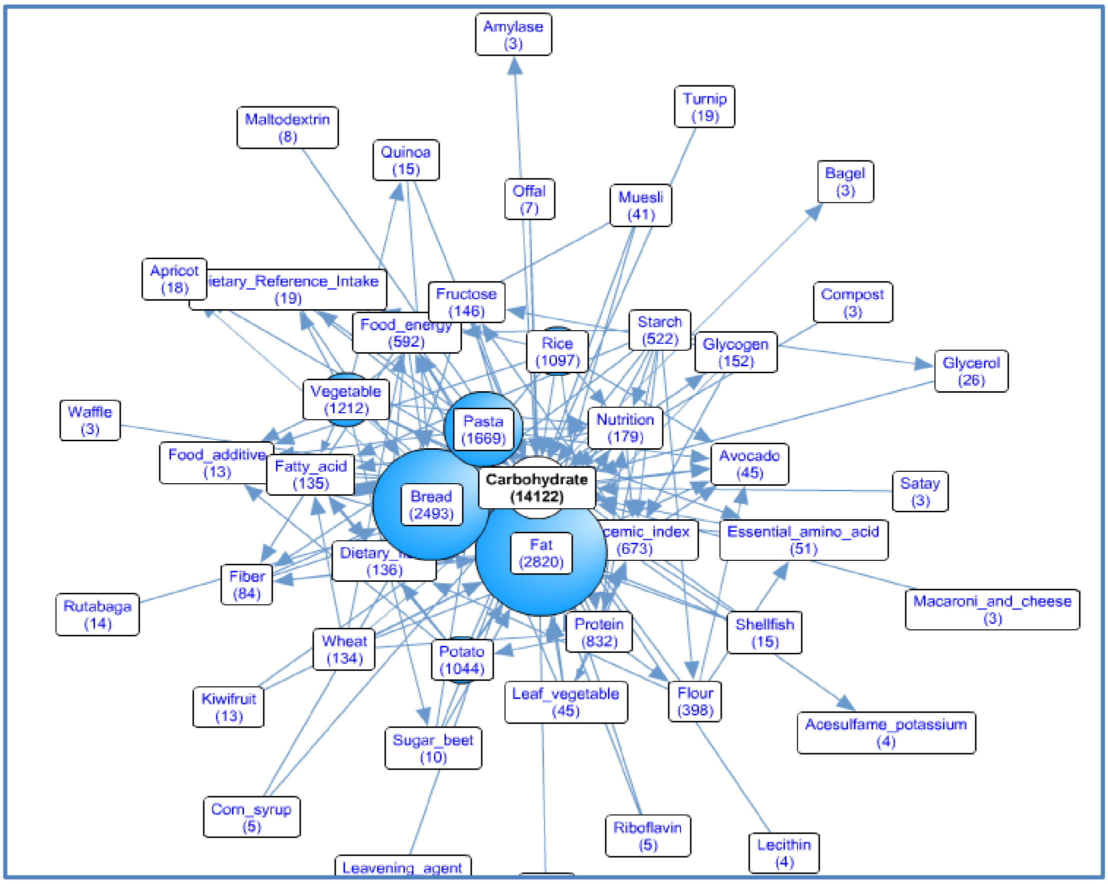


Figure 5-17: Co-occurrence network for Food Category

Even though Figure 5-17 is filled with annotations obtained from the posts, it highlights two advantages of using co-occurrence network to analyse and visualise the extracted annotations. First, using the co-occurrence network allows the analysis and visualisation of a large number of terms and entities, which provide an overview of the discussions in the forums. Second, using the text-mining process, the author was able to simultaneously measure the discussions within different categories, such as food types. Such an endeavour would be difficult and costly using traditional research methods.

Within the "Food and Nutrition" co-occurrence network, several clusters can be observed using SNA analysis. These were identified as *Fruit*, *Vegetables*, *Meat and Sea Food*.

An advantage of using co-occurrence network to visualise the relationships between annotations and terms within the forum, is the ability to *focus* on certain aspects and relations within the network. An example with the LCHF diet is described next.

5.2.3.3 "LCHF diet"-centred Co-occurrence Network

This analysis assess patients experience with one type of diet known as "Low Carbohydrate, High Fat diet" or LCHF. Through the tag cloud analysis, it was observed that this diet has been discussed in the forum using the term "low-carb diet" or "low carb, high fat diet" since 2008, when it first appeared in the tag cloud. However, the term LCHF was used in 2010 for the first time. The derived co-occurrence network around LCHF is shown in Figure 5-18.

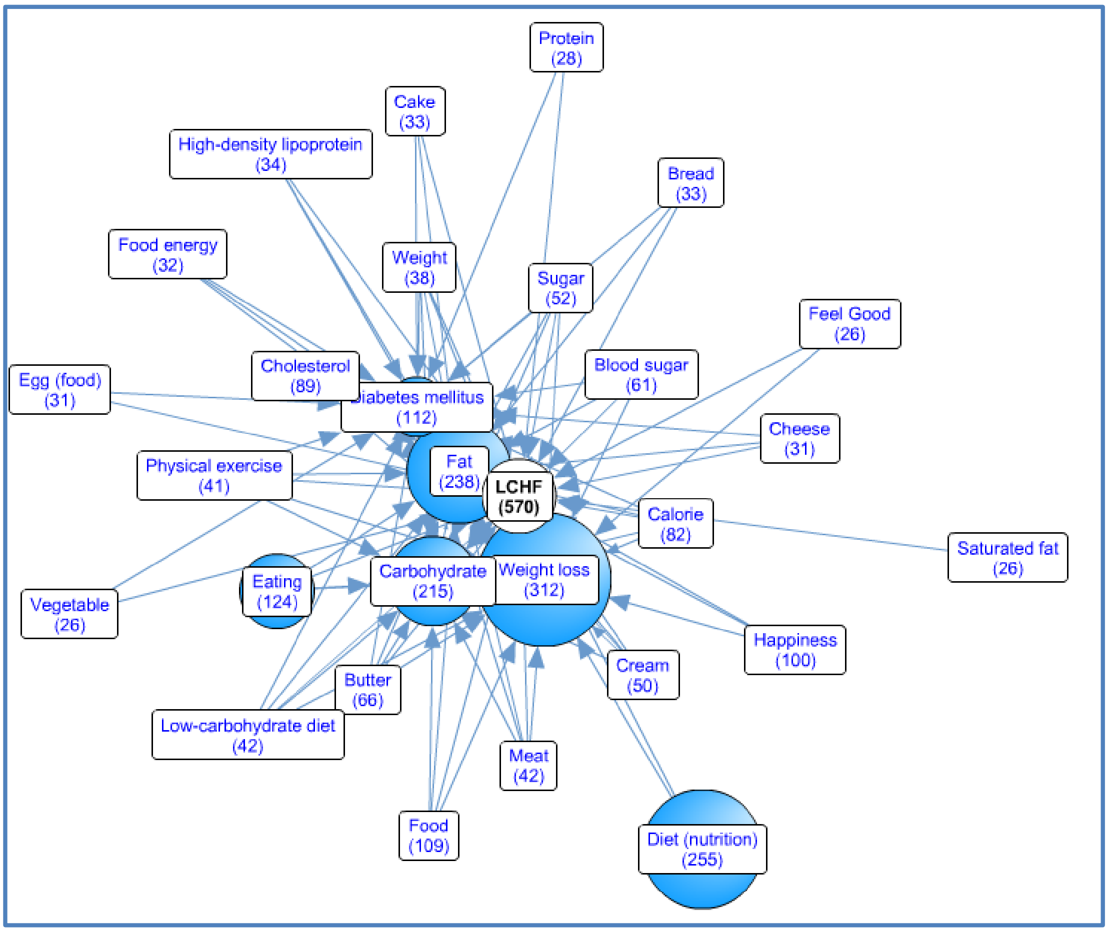


Figure 5-18: LCHF-centred Co-occurrence Network

This co-occurrence network is viewed from LCHF node point of view. Similar to Figure 5-15 and Figure 5-16, in this network, each node represents a tag using a name with a number: The number shows the *popularity* of the tag within the network, i.e., its Centrality value. The edge thickness indicates the absolute weight of the link between the two nodes, while the edge length represents the relevant distance of the nodes based on LCHF perspective.

This network shows the relationship between topics and concepts from the patients' viewpoint regarding diet and associated food. As expected, LCHF is highly associated with "Sugar", "Carbohydrate" and "Fat" annotations. Other interesting associations were found between LCHF and "Weight loss" (70% of posts that include LCHF) as well as *positive* annotations, such as Happiness (30% of the posts that contain LCHF) and "feeling good" (70%). Investigating these links and associations revealed that the LCHF diet has been recommended by more than 75% of the members who tried it and posted about it, although they claimed that their GPs were recommending a "high carb" diet that did not work for them. Another 15% were questioning the "high fat" part of the diet, and were recommending using a different acronym RCIF (Reduced Carbs/Increased Fats) to better reflect what this diet is, while the remaining 10% recommended it for weight loss only.

Other interesting associations with LCHF were "debate", "Saturated Fat" and "Myocardial infraction" annotations. Most members were recommending keeping saturated fats to a minimum to avoid complications, such as cardiovascular diseases. Yet, some members were challenging this with studies that found "evidence of harm" does not appear to be statistically significant.

5.2.3.4 *"Stress"-centred Co-occurrence Network*

Many threads within the diabetes discussion forums contain discussions around stress and its effects on diabetics. Since diabetes is largely a self-managed disease, it requires the patient to have the responsibility for day-to-day management. This leads to high levels of stress amongst diabetics (Pandey et al., 2011). One specific thread in the forum specifically discussed the relationship between blood glucose level and stress, and how music therapy can help patients manage the high levels of stress and, consequently, keeping the blood glucose levels down. Analysing this thread using the derived text mining method results in the co-occurrence network shown in Figure 5-19.

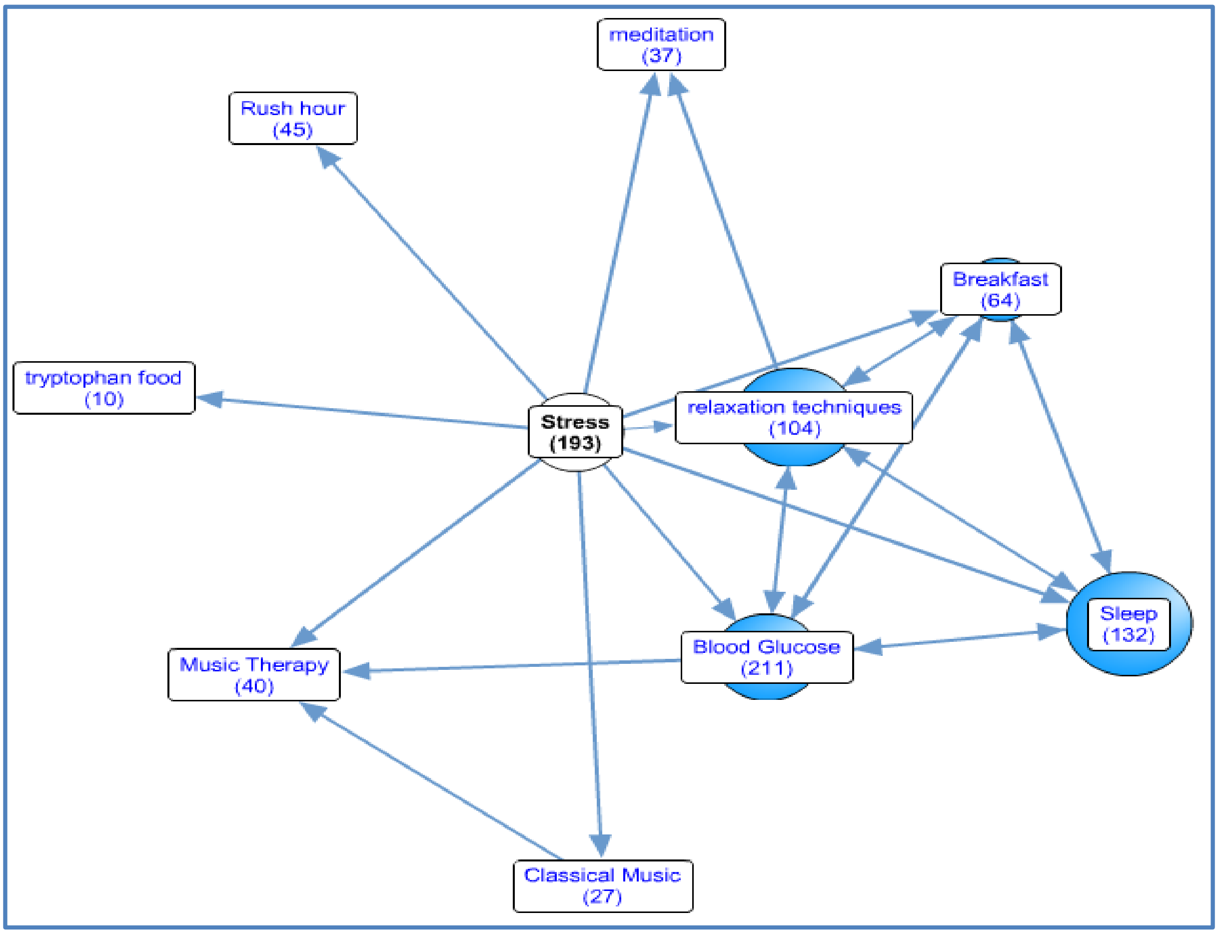


Figure 5-19: Stress-Centred Co-occurrence Network

This co-occurrence network is viewed from ***Stress*** node point of view. This network shows the relationship between topics and concepts from the patients' perspective, in regards to the relationship between stress, blood glucose levels, and different techniques to manage the stress. Music therapy has been identified by the patients as one of the useful relaxation techniques, with a positive effect on the stress, anxiety and blood sugar levels. Certain phrases were associated with music, such as "stress buster" and "healing therapy". Some of the patients have discussed different types of music and how they can affect their stress levels, with classical music emerging as one of the most positive types. In fact, there appears to be new businesses that sell music "tailored for diabetes patients". Patients can also download "Music Therapy for Diabetes" tracks from iTunes (Figure 5-20).

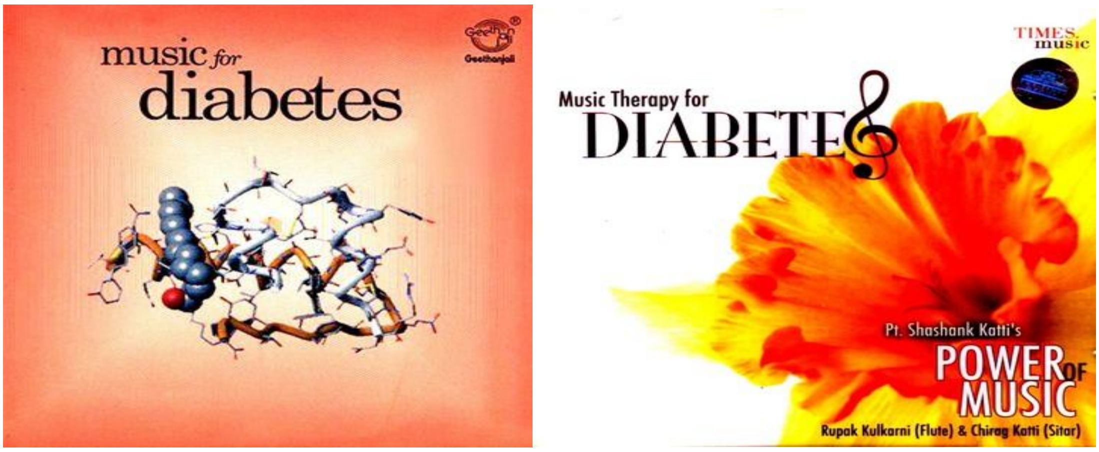


Figure 5-20: Exmples of Music Tailored for diabetes patients

5.3 Discussion

This chapter has presented a text mining system to "*listen*" to online patients' continuing discussions on the diabetes online forum. Semantic annotation based on a hybrid system of domain ontology and a general knowledgebase has been employed in order to identify the main topics and discussion themes within the diabetes forums.

A selection of patterns and insights were found from the analysis of the user-generated content within diabetes forums using the derived text mining method. For example, "There is a strong relation between stress levels, blood pressure and music", "LCHF diet helps you lose weight" and "Chana dal (or Chickpea dahl) keeps glucose readings steady". Some of these insights and patterns have been upheld by research studies in the area. For example, "Music therapy has a positive relation with reducing stress in diabetics" was tested in a feasibility study by Mandel et al. (2013). They conducted a trial with 199 patients with typeI, typeII and prediabetes. The study results support the integrating of music therapy with diabetes self-management as it found a positive relationship between the use of music therapy and lower blood pressure in the patients. The extracted music therapy hypothesis from the forums was originally shared with diabetes researchers from Warwick Medical School in 2012. They were not aware of the use of music therapy for diabetes patients and they were not aware of new businesses that sell music tailored for diabetics.

Using the text from the forum to ascertain and discover the topics and themes of discussions in the studied domain (diabetes) have yielded a preliminary understanding of the users and their interests and concerns. As a result of the analysis, seven main discussion themes (categories) in the diabetes domain were discovered and the relationships were investigated inter- and intra-category.

Grouping the extracted terms and annotations into distinctive categories is one benefit of the proposed text mining method. For instance, in Food and Nutrition category, more than 420 terms were identified, collected and grouped in various clusters according to their semantic annotation subject and type. These sub-categories, such as *Vegetables*, *Fruit*, and *Meat and Fish* gave the posts in the forum extra structure. This structure can facilitate the analysis of the forum as it allows for "zooming in" on a particular sub-category to investigate in depth the relationships between the terms in these sub-categories and with terms in other categories. Another example is in the "Symptoms" category, where the text mining method identified "weight loss", "blurred vision", "fatigue", "Polydipsia", and "increased urination" as symptoms of diabetes. These are the main signs of type I and type II diabetes found in National Health Service (NHS) diabetes guidelines.

The diabetes categories generated could serve as a schema or a high-level hierarchy for existing diabetes forums or for other diabetes portals directed at diabetes patients. Since these categories are extracted from the user-generated content, they reflect the main concerns and interests of the users in the diabetes community. Therefore, organising a forum or a portal around these topics could help users find what they are looking for quickly, and enable them to stay up-to-date with new topics in a specific category through subscribing to a notification system.

The categories and their associated terms can be used to enhance and enrich the derived diabetes domain ontology, or other medical ontologies such as OpenGalen and terminology directories such as SNOMED CT. As the results of co-occurrence network analysis provide both terms and their relationships from the patients' perspective, the outcome of co-occurrence network analysis can be used to improve the navigation of the forums through integrating the results into the diabetes ontology.

The enhanced ontology can then improve the analysis of search queries and could make it easier for the users to find quickly what they are looking for.

The output of co-occurrence network analysis depends on the annotations obtained from the semantic annotation and topic identification. Therefore, it might be sensitive to changes and *uncertainty* in its input. Hence, sensitivity analysis checks were carried out to test the robustness of the model.

5.3.1 Sensitivity Analysis of the Co-occurrence Network

The robustness of the network model outcome was examined by using sensitivity analysis. This analysis of the co-occurrence network examines the uncertainty in its input and its effect on the output network. In this research, the uncertainty in the input may stem from three characteristics:

i) The number of posts mined from the forum (Data Volume),

ii) Users' Participation Behaviour, or

iii) The length of the posts.

5.3.1.1 Sensitivity of the network to Data Volume

This test examines the effect of forum size on the co-occurrence network outcome. The diabetes forum analysed had more than 239,000 posts that covered the period from January 2007 to March 2013. However, it might be possible that other forums in other domains have less data/posts to apply the proposed analysis process. Hence, it is important to test the output sensitivity to the dataset size. Therefore, a cross-correlation analysis was carried out between the co-occurrence network matrix produced by analysing the forum posts and the co-occurrence network matrices produced by randomly choosing half, quarter and one eighth of the posts. As the correlation required is between two dyadic networks, a very basic approach could be to use techniques such as Ordinary Least Squares (OLS) regression or Quadratic Assignment Procedure (QAP) regression (Carrington et al., 2005). However, the variables in the studied networks are not independently distributed, which violates the assumptions of OLS regression (Nagpaul, 2003). Therefore, QAP is going to be used for all sensitivity tests.

QAP provides a calculation for the Pearson's coefficient and simple matching coefficient between two data matrices. It is commonly used in social network analysis (SNA) to observe changes in social networks over a period of time (Hanneman and Riddle, 2005b). Krackhardt (1988) proved that QAP is superior to other techniques, especially OLS for calculating the correlation between dyadic networks.

Table 5-13 shows the QAP correlation values obtained for comparing the co-occurrence network generated from all the posts, with the ones generated by randomly choosing half, quarter and one eighth of the posts respectively, where "r" is the correlation coefficient. From this table, it was observed that the derived co-occurrence networks are similar to the original network (full dataset) even with only one eighth of the posts (Less than 30,000 messages, (r = 0.842, pseudo-p<0.001). Pseudo-P value is an expression of statistical significance. If it is less than "0.05", the correlation would be accepted as statistically significant (Haunschild and Miner, 1997).

Therefore, it can be concluded that number of posts has low or no effect on the quality of the co-occurrence network and the observed insights derived from analysing the posts in the diabetes forum. However, it is observed that for more in-depth analysis of the connections between specific concepts and terms related to them, the output might be more sensitive to the size of the forum. Several terms and entities with low appearance frequency, such as "Hepatitis", have been more sensitive in regards to the size of the sample, which is to be expected. These concepts are not as important due to their low frequency appearance in the discussions, and can be filtered out while analysing the discussions.

Table 5-13: Comparison of the co-occurrence networks based to the dataset size

Data Size	Correlation Coefficient (r) Value
Half of the posts	0.951
Quarter of the posts	0.879
Eighth of the posts	0.842

5.3.1.2 Sensitivity of the network to Users' Participation

This sensitivity test explores the effect of forum members' activity on the robustness of the co-occurrence model. While analysing the diabetes forum, all the posts were collected within a period of six years. Therefore, the analysis of the forum and the corresponding co-occurrence network should essentially echo the discussion within the forum. However, one possible issue when analysing the forums' posts is the variation in contribution by different members on the forum, where few users are more active than the rest and they generate most of the forum's content. In the diabetes forum case, 10% of users have contributed more than 80% of the posts and 90% of the users have only written once.

In order to test the effect different users have on the quality of the co-occurrence network, a cross-correlation between the co-occurrence network matrix of active users and the co-occurrence network matrix of less active users has been carried out. The definition of active users in our case is the ones with more than ten posts in the forum (10% of the users who generated 82% of the content). The other group of users with less than ten posts forms 90% of the overall users and generated the remaining 18% of the posts. QAP correlation was calculated between the two matrices that represent the co-occurrence networks, and it was relatively high with r =0.923, pseudo-p <0001. This high value suggests that for this particular forum, the active 10% of forum members can represent the overall community.

5.3.1.3 Sensitivity of the network to Length of Posts

In the forum, the post length might be connected with the post's type (such as a question vs. answer) or the user's type, such as novice or expert. In order to test the sensitivity of the co-occurrence network to the length of posts, two groups based on the length of the posts, Long posts and short posts were created. Long posts are defined as posts that contain more than the median number of sentences in a post, which in the case of the Diabetes forum was 9 sentences. Naturally, longer posts have a bigger chance for concepts and terms to co-occur. In this dataset, 70% of the long posts accounted for 60% of the co-occurrences in the network matrix. Surprisingly, the cross-correlation between the two matrices based on long and short posts is high (0.86, pseudo- p <0001), which suggests that the co-occurrence network is robust to different posts lengths.

5.4 Generalisability Analysis

The text-mining process applied in this research was derived for a specific case – diabetes discussion forums. Hence, it is useful to analyse the derived method for generalisability, i.e. whether the text mining method could address other forums with different structure and other domains with different levels of information, and be utilised by other researchers.

5.4.1 Generalisation of the Case Study Results

Changing the case study or the domain might cause some differences compared to the investigated case. External validity examines the ability of the research results to be applied for other forums or other domains (Collis and Hussey, 2003).

The case study of diabetes examined in this research showed two important characteristics that could aid the generalisation to other domains: Domain Independence, and flexibility.

5.4.1.1 Domain Independence of Semantic Text Mining Method

Domain independence means that the methods and techniques of the derived text mining method can be used in other domains and cases. The text mining method demonstrates domain independence in both its semantic annotation and the visual analytics components. As the latter is domain independent by its nature, this section focuses on the domain independence only within the semantic annotation phase.

The semantic annotation and topic identification phase depends on two knowledge sources to perform its task: the domain ontology and DBpedia knowledgebase. Therefore, their domain independence characteristics are discussed.

5.4.1.1.1 Domain Independence in the Ontology Building Process

The ontology building process exhibits domain independence through two main features: Knowledgebase coverage and the semantic relationships obtained. Knowledgebase coverage and semantic relationships contribute simultaneously to the ontology building process consistency, which is a valuable feature of the text mining method.

I. The ontology corpus building process has used the OneLook Reverse Dictionary as the knowledge source (discussed in section 3.1.3). The OneLook Reverse Dictionary is not restricted to a specific domain as it indexes sources from all the domains and provides semantic related concepts in most of them. The coverage of the indexed data sources ranges from general ones, such as Wikipedia, to specialised ones such as the “Minerals and Metals” Dictionary. OneLook as a search engine also has an advantage from the *recentness of information* viewpoint. It updates and indexes the sources and dictionaries on a regular basis. In 2014, its index contained 19,633,000 words from more than 1060 sources.

II. Choosing semantic relatedness as the relationship type for generating terms from the seeding keywords (discussed in section 3.1.2) is equally useful in other cases and domains. This type of relationship is not restricted to a particular domain as long as the domain has suitable coverage of its content. For example, semantic relatedness was used and tested by Ma et al. (2014) to derive an Engineering domain ontology for resource matching.

III. Using this characteristic, and if the domain has enough coverage, the ontology corpus building process might be able to obtain large number of domain terms with similar structure to the diabetes ontology examined in this research. However, the distribution of concepts in the three zones may vary depending on the characteristics of the domain itself or the coverage of the domain within the OneLook index. For example, the distribution of concepts within the derived diabetes ontology was different compared to the feasibility study output, although they all had a three-zone network structure.

Some limitation might be experienced using the ontology building process when applied to some domains. Obtaining domain-specific terms from OneLook depends on the general interpretation of the domain in the indexed sources of OneLook. However, in some cases, this general perception of the domain might not be suitable for other applications that have a specific requirement. For instance, a highly specialised and customised medical application might need a tailored thesaurus-based medical ontology in order to reflect its own interpretation of this medical domain and the relationships between its concepts. In this case, the general interpretation of the knowledge provided by OneLook might not be appropriate for an application's particular requirements and it would not be suitable for deriving the required ontology from an end-user application's viewpoint. For example, a derived ontology from a "global warming is real" viewpoint by experts will differ from a "global warming is false" viewpoint. However, it could be argued since the methodology probably reflects both the positive and negative viewpoints in proportion to the amount of material indexed from these viewpoints in online sources, it is a valid interpretation.

Based on the output of the case study and this discussion, it can be reasonably suggested that it is possible to produce an ontology using this process for different domains, even though there might be some difficulties while producing tailored ontology for particular applications.

5.4.1.1.2 Domain independence in the DBpedia Knowledgebase

The semantic annotation process uses a hybrid approach for tagging the online discussion forum. The domain ontology and DBpedia knowledgebase are used to annotate the posts, and identify the topics in the forum's threads.

Since it relies on DBpedia knowledgebase, DBpedia Spotlight could be used to annotate text from different domains as long as there is enough content coverage of that domain in Wikipedia and DBpedia. As discussed in section 2.4.2, DBpedia is considered one of the biggest multi-domain knowledgebase, and interlinks with other ontologies, such as Freebase and WordNet. It covers multiple domains and evolves as Wikipedia changes. This gives DBpedia an advantage when new terms appear in the domains, since it automatically updates its entries using Wikipedia pages. In October 2014, 4.22 million concepts were classified in its English version ontology. These concepts cover many domains including medical, ICT, literature and others. It can be assumed that the broad, deep and up-to-date coverage of DBpedia allows the semantic annotation to be used in different domains.

However, as discussed in section 5.4.1.1.1, the general knowledge interpretation offered by Wikipedia and DBpedia might not be suitable for particular applications. An assessment of the domain coverage needs to be performed through a feasibility study in order to test the applicability of DBpedia Spotlight for the studied domain.

5.4.1.2 Flexibility in Semantic Text Mining Method

The text mining method offers a significant level of flexibility in its phases. In this research, flexibility is the ability to easily customise the process according to the application requirements. The flexibility exhibits in all phases of this process, i.e. the ontology building process, semantic annotation and visual analytics.

I. The building process of the domain provides flexible result display. This is demonstrated in rich internal structure as a network, which can be highlighted in the top zone of the ontology. This feature allows the researchers to focus the output around specific terms or clusters.

   In addition, the relationships obtained between the terms in the derived ontology could be customised by the researchers to suit the application requirements. The customisation can be achieved, for instance, by restricting the relationships weight.

For example, Figure 5-21 shows a “typeII diabetes” network, where the relationship weight was limited to $f_{RD} \leq 1.86$ and $f_{RD} \leq 2$ respectively.

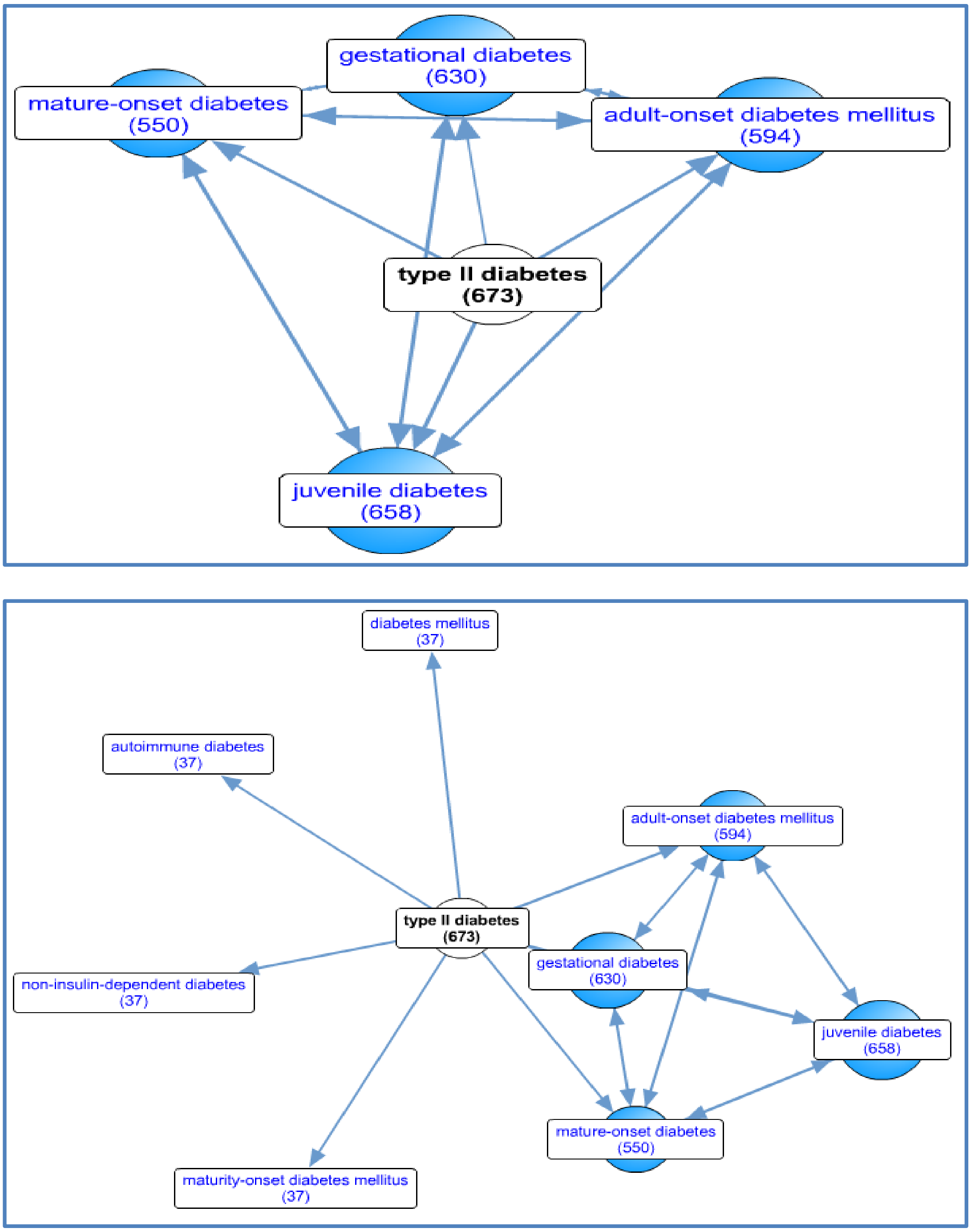


Figure 5-21: TypeII diabetes customised network - $f_{RD} \leq 1.86$ and $f_{RD} \leq 2$

II. The semantic annotation allows the configuration of each step of the annotation process according to the different requirements. This flexibility is due to the use of the domain ontology and the DBpedia Spotlight for semantic annotation.

Moreover, the ability to use SPARQL queries with DBpedia Spotlight provides more flexibility to the system. For example, the researcher or the system developer can decide to use only concepts and annotations that belong to particular historical period or related to a particular country.

DBpedia Spotlight output can be obtained in different formats, such as RDF, JSON, or XML. In these formats, each annotation within the analysed text is associated with the text part where it was identified. Figure 5-22 shows an example of DBpedia Spotlight XML formats when annotating a sample text from the diabetes forums.

This allows easier integration with existing systems, where developers and researchers can use a specific output format according to their applications required format.


```
<Resource
types="DBpedia:Plant,DBpedia:Eukaryote,DBpedia:Species,Freebase:/food/ingredie
nt,Freebase:/food" support="940"
percentageOfSecondRank="0.2182506749758442"
similarityScore="0.22970135509967804" offset="119" surfaceForm="orange"
URI="http://dbpedia.org/resource/Orange_%28fruit%29"/>

<Resource
types="Freebase:/book/book_subject,Freebase:/book,Freebase:/chemistry/chemic
al_classification,Freebase:/chemistry,Freebase:/food/ingredient,Freebase:/food,Fre
ebase:/food/food,Freebase:/food/nutrient,DBpedia:TopicalConcept"
support="2568" percentageOfSecondRank="0.4118355173206223"
similarityScore="0.17656682431697845" offset="229" surfaceForm="sugar"
URI="http://dbpedia.org/resource/Sugar"/><Resource types="" support="107"
percentageOfSecondRank="-1.0" similarityScore="0.15688271820545197"
offset="240" surfaceForm="sweetener"
URI="http://dbpedia.org/resource/Sugar_substitute"/>
```


Figure 5-22: DBpedia Spotlight XML format

III. The Visualisation methods utilised, dynamic tag cloud generation and co-occurrence network, are able to be configured and can have a high degree of flexibility. As discussed in 5.2.2, a dynamic tag cloud can be configured to show trends within a particular time period (2008-2012 in diabetes forums case).

Co-occurrence can be configured to visualise part of the data, as discussed in sections 5.2.3, such as a "Food and Nutrition" category and the relationships between the entities within that category.

In summary, the domain independence and flexibility characteristics 5.4.1.2of the text mining method created demonstrate its potential ability to be generalised to other domains.

5.5 Generalisability for Researchers and Text Analysts

The generalisability of the semantic text mining method derived should involve whether it can be used by other researchers easily, quickly and in a cost-effective manner (Ward-Schofield, 1993). There follows a discussion of these process's characteristics.

I. The different components within the proposed semantic text mining method demonstrate cost-effective analysis tasks. As discussed in section 2.3.1, ontology generation methodologies suffer from the reliance on experts in the domain studied, which is a significant obstacle to the reuse and customisation of existing ontologies for different applications. The method derived has reduced the reliance on domain experts significantly.

In addition, the semantic annotation used in this method is an economic annotation process since it is based on DBpedia Spotlight. It can be accessed either as a free Web Service or locally as a standalone installation. Using the Web Service, custom applications can be easily built to re-use the system and its functions. For example, both SOAP and RESTful web services can be used to access the annotating and disambiguation functionalities provided by DBpedia Spotlight. An alternative to using the Web Service for annotation and disambiguation is installing and using the system locally within a company or a research laboratory. This tackles issues regarding the analysis of private sensitive data without sending it through a Web Service. The local installation can be obtained for free since DBpedia Spotlight is open source and its installation package can be obtained online.

II. The semantic text mining method derived can be conducted relatively quickly compared to traditional methods. When a large corpus is to be analysed it is suggested that an initial feasibility study be attempted. An initial feasibility study can be performed quickly and are capable of producing useful results in less than a week. These estimates came from the experience of this research with three feasibility studies in "Cardiovascular", "Cancer" and "Diabetes" domains. This gives text analysts more time to spend on trying different configurations and customisation that suit their application requirements.

III. The generalisability of the process is also enhanced through the ability to distribute tasks and enable collaboration between different researchers. For example, the semantic annotation phase can be run separately by different researchers on different parts of the data. This provides a parallel textual analysis capability, which enhances the performance when analysing vast amount of text and allows for system scalability. This collaborative and parallel capability can further lead to a cloud-based architecture for the semantic text-mining process.

Consequently, more complex text analysis tasks can be performed in a fast and economic manner.

5.6 Summary

This chapter has sought to present the results of the semantic text mining method in the diabetes case study and evaluate the results to demonstrate that this method is capable of useful analyse through improved visualisation and structuring of the text in online discussion forums. This method was capable of identifying the concepts and topics in the text, the strength of relationships between them, and drive the process of identifying emergent knowledge through a novel data visualisation.

The text mining method derived has primarily been devised and tested for the diabetes discussion forums. However, it appears to be extendable to other domains from an external viewpoint. It can also be used by other researchers and text analysts for analysing online discussion forums.

Chapter 6: Conclusion

The research presented in this thesis proposes new methods to help identify the "known unknowns" and "unknown, unknowns" in a domain. This phrase was made (in)famous by Donald Rumsfeld in 2002, but does highlight a general problem. Data mining techniques have provided solutions to some of these questions, but access to sufficient quantity and quality of data has nearly always been a major problem.

Now with the data being captured on the internet doubling in size every two years (Turner et al., 2014), we have new sources of information that may contain better answers to the questions we are asking. Currently, most of this data, generated by users, is in text form. This provides a rich source of information that can aid understanding in key domains, such as the medical domain. However, the information contained in this data is mostly unstructured and difficult to analyse.

The Semantic Web has been a key area for research in recent years, looking for ways to structure and analyse this data through enriching and annotating in a machine-understandable fashion so that automated methods can help us find the known and the unknown, unknowns.

The example field chosen in this research was diabetes because it is increasingly one of the most common diseases that affects the population of the World. The successful management of this disease requires constant and dedicated efforts from both patients and healthcare professionals. With this chronic disease, it is essential for patients to become their own health managers and treatment experts. Therefore, they are actively searching for information online about diabetes for different purposes, such as treatment, diet and support (Greene et al., 2011). That also encourages patients to participate and engage in online forums and communities (Zhang et al., 2013).

To assist communication between patients and medical professionals and to help uncover interesting knowledge, it is important to analyse the discussions of sufferers on online diabetes forums to feedback the topics under discussion and the results from the many experiments that they conduct and discuss.

In this research, a novel hybrid method to help analysing the text in discussion forums to tap into the vast amount of data regarding patients' thoughts, experiences and personal experimentation has been devised. In general, this method helps investigators to explore user-generated content and identify interesting patterns and emergent knowledge. To achieve this objective, this research has sought to answer the research question:

> "How can we identify and display the concepts and topics being discussed, the strength of the relationships between them and the emergent knowledge in online discussion forums?"

Based on this question, three main sub-questions were identified:

1. How can we identify the concepts in a particular domain and the relationships between them? This involves both at an expert level but also at a novice level and linking these two.
2. How can we understand the emergent knowledge behind text conversations?
3. How can we present the living nature of an online discussion forums with themes dying, being reborn or new ones emerging?

These questions were addressed through introducing a novel semantic text mining method for analysing text in the discussion forums, and experiments were carried out using diabetes forums as a case study. In summary, the main conclusions drawn from this research are:

1. Text mining systems would benefit from including a domain knowledge source that can connect the terminology used by online users and the one used by domain experts. This also helps bridging the semantic gap observed when using traditional text mining methods, and enable a better understanding of the text. Existing domain ontologies and lexicons methodologies depend on the domain experts in order to obtain domain-focused concepts and relationships. This presents a challenge in employing these knowledge sources since domain experts are costly to employ and often do not agree on what to include in the knowledge source.
2. A new semantic text mining method is suggested by this research to tackle the challenging task of analysing and visualising discussions on online forums. The process has demonstrated the ability to build a domain knowledgebase automatically for the text mining method, with little need for domain experts. This derived ontology can have the breadth and depth of coverage required for text mining tasks, with a rich relationship structure that enhances the semantic annotation of text.
3. A hybrid semantic annotation system was proposed in order to annotate and analyse the text. This system employs domain ontology and general knowledgebase (DBpedia) to capture and analyse topics in the text. Using this hybrid system in the diabetes forum case study, it was possible to semantically annotate the posts and identify different topics discussed by forum users. Then, a visualisation text mining techniques of dynamic tag clouds and co-occurrence network were used to carry out pattern discovery and analysis.

Using these methods, relationships between interesting categories and concepts were identified and explored in order to discover new knowledge. Examples include the identification of music therapy as a main topic and the emerging agreement about diet choices among diabetes sufferers that had not been observed by the experts.

4. The text mining method derived should and seems to have performed better in semantic annotation of the text compared to other systems.
5. The text mining method has the potential to benefit the analysis of forums in other domains by other researchers easily, quickly and in a cost-effective manner. This process could release valuable feedback to experts in fields where online discussion forum users are conducting informal experiments and commenting on the results. Particularly for health forums this has great potential for broadening the understanding of experts. Partially from knowing what patients are really doing as opposed to what they tell their doctors.

6.1 Contribution to Knowledge

This thesis contributes to text mining literature and its related fields in the following ways:

1. A novel semantic text mining method was proposed to analyse and visualise text extracted from online discussion forums. This process enables fast analysis and visualisation of concepts in the text and requires little contribution from domain experts for building the domain knowledge structure essential to enrich and analyse the text.
2. As part of this text analytics system, the research contributes a novel domain ontology building process that is suitable for analysing text in online forums.

The method derived provides flexible width and depth of domain coverage as well as linking the generated domain concepts and terms to rich external knowledgebase in order to annotate and enrich the text. The derived ontology has a rich network structure which facilitates investigation in a variety of ways, and resembles a faceted ontology. Generating a rich structure of internal relationships is often limited in ontologies that rely on domain experts for obtaining the concepts and their relationships as a result of the *cost* of their time. The derived network structure also provides weighted and directed relationships within the ontology. The number of internal relationships can be increased and the strength of the relationship quantified compared to existing comparable ontology.

3. This novel text analytics approach contributes a semantic annotation and topic identification process using a hybrid system of domain ontology and DBpedia knowledgebase without the need for machine learning methods or training data.
4. The text analytics system includes trend analysis component using Dynamic Tag Clouds: This study contributes a dynamic tag cloud and trend analysis for both terms and annotation based on information theory concept of entropy. This visualisation combines a tag cloud importance curve with the generated tag clouds in order to demonstrate the evolution of topics and discussions in the online discussion forum. The importance curve is constructed based on the mutual information measure between time-based tag clouds.
5. Co-occurrence Network-based Analysis: This study also contributes a co-occurrence network analysis and visualisation for the extracted entities and categories from the semantic annotation phase. This network can be

customised and parameterised to show relationships between concepts and annotations within a category or between selected categories. The ability of the network to show the relationships between categories and concepts is a main advantage for identifying and presenting new patterns in the analysed data.

6. Modular Design of the Text Mining Method: The derived semantic text mining method has a modular design where each stage has a well-defined input and output. This design provides the ability to re-use the components of the process on their own for other applications.

6.2 Research Limitations

There are two main limitations in this study that could be acknowledged – The process representativeness and the reliance on online tools.

6.2.1 Text Mining Method Representativeness

The semantic text mining method has been employed to analyse the text data obtained from a diabetes forum. Pilot studies in other domains were carried out, such as “Cardiovascular diseases” and “Cancer”. However, these cases might not be sufficiently representative of the general situation for the method to be employed in other domains. Further assessment is required regarding the applicability of the text mining method to other non-health related domains. This might be done through examining the quantity and quality of information available about the studied domain within the knowledge sources that the text mining method used. However, as discussed in chapter five, this process appears generalizable to other domains. Any limitations do not arise from the method itself but from the data sources it makes use of. The main data sources, OneLook, itself indexes information from a wide range of others, and DBpedia (derived from Wikipedia) is one of the world’s largest sources, it is believed that the method may well be as robust as we can currently achieve. This is discussed further in the next section.

6.2.2 Dependence on External Online Tools

Obtaining terms and concepts from both OneLook and DBpedia Spotlight in order to form the domain knowledge might be considered as a limitation to the derived ontology and the text mining method.

OneLook and DBpedia index and archive public information online. OneLook is a search engine that indexes definition entries and dictionary websites across the internet as well as other sources such as Wikipedia articles. Whereas DBpedia is an online knowledgebase created by automatically extracting structured information from Wikipedia. Therefore, the reliability of the information retrieved from these sources might be questionable, especially the Wikipedia pages. From the ontology building perspective, the quantity and quality of available indexed sources online around a particular domain might affect the snowball sampling rounds while retrieving the terms and concepts from OneLook.

In addition, the reliance on an external online tool contains a risk regarding the discontinuity of the available services needed by this research. OneLook and DBpedia Spotlight might be discontinued at some time in the future by their owners or developers. In that case, an assessment for an alternative knowledge source has to be carried out to replace the discontinued services. The text mining method in this research has been built in a modular way, so it is possible to replace the knowledge sources with equivalent ones as long as they provide similar functionality. Examples of equivalent annotation tools that could replace DBpedia Spotlight are Zemanta and OpenCalais.

6.3 Future Research Opportunities

The analysis of the forums discussions reflects the views of its members, which may or may not reflect the views of the wider population of diabetes patients. The high number of users (more than 116,000 members) within the forum might give an indication, but a future study could investigate this issue in a formal way. The automatic text mining method minimises the effect of recall bias and promotes commonalities within the forum. Yet, patients' views on the forum might still be biased as they aim to promote their own views within the forum.

In fact, there exists a lot of pressure exists on firms that operate in a highly competitive market to analyse their market positions after introducing a new product or deploying a new service (Kumar, 2012). The ability of the derived text mining method to visualise the trends and patterns as a time-based dynamic tag cloud, or co-occurrence network provides an interesting possibility of analysing the discussions and trends before and after a specific event occurs. This might be used by pharmaceutical companies, for example, to analyse the market before and after introducing a new drug. It can also be used to analyse the reactions of customers to new products in competitive markets, such as the introduction of iPhone6 and comparing it with the introduction of other products, such as Nexus 6, in the same period.

This process can also be applied in well-established non-medical domains, where the semantic text mining method can unveil interesting patterns and trends from their discussion forums, as well as a market overview (known as a perceptual mapping) of the domain. An example of such domain might be the car industry or sport clothing brands. Automatic *surveillance* of market structure and obtaining market insights is an existing research problem (Feldman et al., 2010). The text mining approach presented in this study could be modified to tackle this research issue. The reviews of different products and brands in the domain can be mined and comparison between the products and their features can be shown as a co-occurrence network. A dynamic tag cloud can also be used to visualise the trends within the industry. This can be considered as a “market structure” analysis based on the customers reviews, which can be downloaded from multiple sources.

Future studies can also investigate the generalisability of the semantic text mining approach in emerging domains rather than established ones in order to provide preliminary understanding of the needs and issues faced by customers in those domains, such as 3D printers or wearable devices (e.g. Google Glass). This could serve as an automatic process for customer feedback to manufacturers of these technologies in order to improve the designs and address the problems as they appear. Another interesting area of future research is an investigation of the suitability of the derived text mining approach to other types of social media, such as blogs, or to analyse news websites to reveal interesting connections, such as the relationship between news in general (or financial in particular) and the stock market index.

The generalisability of the text mining approach to be applied to analyse forums written in languages other than English is a challenging research problem. This aspect has not been tested in this study and could be an interesting future research.

In summary, the semantic text mining method derived in this research could provide a key tool for examining the patterns and trends hidden in the vast amount of rich unstructured text available on social forums, and help us answer the known unknowns, and possibly also the unknown, unknowns!

References

Aas, K. & Eikvil, L. (1999). Text categorisation: A survey. *Raport NR 941*.

Abbar, S., Amer-Yahia, S., Indyk, P. & Mahabadi, S. (2013). Real-time recommendation of diverse related articles. Proceedings of the 22nd international conference on World Wide Web. International World Wide Web Conferences Steering Committee. pp. 1-12.

Abel, F., Gao, Q., Houben, G.-J. & Tao, K. (2011). Semantic enrichment of twitter posts for user profile construction on the social web. *In:* ANTONIOU, G., GROBELNIK, M., SIMPERL, E., PARSIA, B., PLEXOUSAKIS, D., LEENHEER, P. D. & PAN, J. (eds.) *The Semanic Web: Research and Applications.* Springer. pp. 375-389.

Abulaish, M. & Dey, L. (2007). Biological relation extraction and query answering from medline abstracts using ontology-based text mining. *Data & Knowledge Engineering*. 61(2)**.** pp. 228-262.

Agarwal, N., Liu, H., Tang, L. & Yu, P. S. (2008). Identifying the influential bloggers in a community. *Proceedings of the 2008 International Conference on Web Search and Data Mining*. Palo Alto, California, USA: ACM.

Aggarwal, C. C. & Zhai, C. (2012a). An introduction to text mining. *In: Mining text data.* Springer. pp. 1-10.

Aggarwal, C. C. & Zhai, C. (2012b). *Mining text data*. Springer.

Agichtein, E., Castillo, C., Donato, D., Gionis, A. & Mishne, G. (2008). Finding high-quality content in social media. *Proceedings of the 2008 International Conference on Web Search and Data Mining.* Palo Alto, California, USA: ACM.

Ahmadabadi, M. N., Imanipour, A., Araabi, B. N., Asadpour, M. & Siegwart, R. (2006). Knowledge-based extraction of area of expertise for cooperation in learning.

Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on. IEEE. pp. 3700-3705.

Anand, S. S., Bell, D. A. & Hughes, J. G. (1995). The role of domain knowledge in data mining. *Proceedings of the fourth international conference on Information and knowledge management.* Baltimore, Maryland, USA: ACM.

Anderson, J. R. (2005). *Cognitive psychology and its implications.* Macmillan.

Anderson, J. R. & Pirolli, P. L. (1984). Spread of activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* 10(4)**.** pp. 791-798.

Anna, J. (2010). *LOINC and other standards* [Online]. Available from: http://loinc.org/faq/getting-started/loinc-and-other-standards [Accessed: 11/10/2014].

Bakhshi-Raiez, F., de Keizer, N., Cornet, R., Dorrepaal, M., Dongelmans, D. & Jaspers, M. W. (2012). A usability evaluation of a SNOMED CT based compositional interface terminology for intensive care. *International journal of medical informatics.* 81(5)**.** pp. 351-362.

Bakshy, E., Hofman, J. M., Mason, W. A. & Watts, D. J. (2011). Everyone's an influencer: quantifying influence on twitter. *Proceedings of the fourth ACM international conference on Web search and data mining.* Hong Kong, China: ACM.

Banerjee, S., Ramanathan, K. & Gupta, A. (2007). Clustering short texts using wikipedia. *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval.* Amsterdam, The Netherlands: ACM.

Barrios, M. A. & Vilches-Blázquez, L. M. (2010). Is it possible to enrich ontologies with a specialized domain linguistic resource? *Establishing and using ontologies as a basis for terminological and knowledge engineering resources* [Online]. Available:

http://oa.upm.es/6168/1/Barrios.Vilches.definitive_article2.pdf [Accessed 16/03/2012].

Bengoetxea, K., Agirre, E., Nivre, J., Zhang, Y. & Gojenola, K. (2014). On WordNet semantic classes and dependency parsing. 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, Maryland, USA. ACL. pp. 649-655.

Berry, M. W. & Castellanos, M. (2008). Survey of text mining II: Clustering, Classification, and Retrieval. Springer.

BioPortal. (2014). *Ontology of Glucose Metabolism Disorder* [Online]. Available from: http://purl.bioontology.org/ontology/OGMD [Accessed: 06/01/2014].

Bloehdorn, S. & Hotho, A. (2006). Boosting for text classification with semantic features. *In: Advances in Web mining and Web usage Analysis.* Springer. pp. 149-166.

Bodenreider, O. & Burgun, A. (2005). Biomedical ontologies. *In: Medical Informatics.* Springer. pp. 211-236.

Bontcheva, K. & Cunningham, H. (2011). Semantic annotation and retrieval: Manual, semi-automatic and automatic generation. *In:* DOMINGUE, J., FENSEL, D. & HENDLER, J. A. (eds.) *Handbook of Semantic Web Technologies.* Springer.

Bontcheva, K. & Rout, D. (2014). Making sense of social media streams through semantics. *Semantic Web Journal.* 5(5)**.** pp. 373-403.

Brown, S. H., Elkin, P. L., Bauer, B. A., Wahner-Roedler, D., Husser, C. S., Temesgen, Z., Hardenbrook, S. P., Fielstein, E. M. & Rosenbloom, S. T. (2005). SNOMED CT®: utility for a general medical evaluation template. AMIA Annual Symposium Proceedings. American Medical Informatics Association. pp. 101-105.

Budanitsky, A. & Hirst, G. (2001). Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures. Workshop on WordNet and Other Lexical Resources.

Bumgardner, J. (2006). *Building tag clouds in perl and php.* O'Reilly Media, Inc.

Buranarach, M., Supnithi, T., Chalortham, N., Khunthong, V., Varasai, P. & Kawtrakul, A. (2009). A semantic web framework to support knowledge management in chronic disease healthcare. *In: Metadata and semantic research.* Springer. pp. 164-170.

Byron, L. & Wattenberg, M. (2008). Stacked Graphs-Geometry & Aesthetics. *IEEE Trans. Vis. Comput. Graph.* 14(6)**.** pp. 1245-1252.

Carmel, D., Roitman, H. & Zwerdling, N. (2009). Enhancing cluster labeling using wikipedia. Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. ACM. pp. 139-146.

Carmel, D., Uziel, E., Guy, I., Mass, Y. & Roitman, H. (2012). Folksonomy-based term extraction for word cloud generation. *ACM Transactions on Intelligent Systems and Technology (TIST).* 3(4)**.** pp. 60.

Carrington, P. J., Scott, J. & Wasserman, S. (2005). *Models and methods in social network analysis.* Cambridge university press.

Chalortham, N., Buranarach, M. & Supnithi, T. (2009). Ontology Development for Type II Diabetes Mellitus Clinical Support System. Conference Ontology Development for Type II Diabetes Mellitus Clinical Support System.(Year).

Charola, A. & Machchhar, S. (2013). Comparative study on Ontology Based Text Documents Clustering Techniques. *Data Mining and Knowledge Engineering.* 5(12)**.** pp. 426.

Church, K. W. & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational linguistics.* 16(1)**.** pp. 22-29.

Collis, J. & Hussey, R. (2003). Business research: A practical guide for postgraduate and undergraduate students. New York: Palgrave Macmillan.

Cover, T. M. & Thomas, J. A. (2006). *Elements of information theory.* 2nd edition edition. John Wiley & Sons.

Cui, W., Wu, Y., Liu, S., Wei, F., Zhou, M. X. & Qu, H. (2010). Context preserving dynamic word cloud visualization. Pacific Visualization Symposium (PacificVis), 2010 IEEE. IEEE. pp. 121-128.

Deloitte (2014). 2014 Global health care outlook.

Diabetes UK. (2014). *DIABETES PREVALENCE 2013* [Online]. Available from: http://www.diabetes.org.uk/About_us/What-we-say/Statistics/Diabetes-prevalence-2013/ [Accessed: 23/03/2014].

Dou, D., Wang, H. & Liu, H. (2015). Semantic data mining: A survey of ontology-based approaches. IEEE International Conference on Semantic Computing (ICSC). IEEE. pp. 244-251.

Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P. & Tomkins, A. (2007). Visualizing tags over time. *ACM Transactions on the Web (TWEB).* 1(2)**.** pp. 7-16.

EAGLES. (1998). Expert Advisory Group on Language Engineering Standards. Preliminary Recommendations on Semantic Encoding. Available: http://www.ilc.cnr.it/EAGLES96/rep2/node37.html.

Feldman, R., Goldenberg, J. & Netzer, O. (2010). Mine your own business: Market structure surveillance through text mining. *Marketing Science Institute, Special Report.* pp. 10-202.

Feldman, R. & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data.* Cambridge University Press.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry.* pp. 35-41.

Gabrilovich, E. & Markovitch, S. (2005). Feature generation for text categorization using world knowledge. *Proceedings of the 19th International Joint Conference on Artificial Intelligence.* Edinburgh, Scotland: Morgan Kaufmann Publishers Inc.

Ganendran, G., Tran, Q.-N., Ganguly, P. & Ray, P. (2002). An ontology-driven multi-agent approach for healthcare. *HIC 2002: Proceedings: Improving Quality by Lowering Barriers.* pp. 464.

Ghose, A. & Ipeirotis, P. G. (2011). Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics. *Knowledge and Data Engineering, IEEE Transactions on.* 23(10)**.** pp. 1498-1512.

Gjoka, M., Butts, C. T., Kurant, M. & Markopoulou, A. (2011). Multigraph sampling of online social networks. *Selected Areas in Communications, IEEE Journal on.* 29(9)**.** pp. 1893-1905.

Goh, K.-Y., Heng, C.-S. & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Information Systems Research.* 24(1)**.** pp. 88-107.

Gómez-Pérez, A., Fernández-López, M. & Corcho, O. (2004). *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. (Advanced Information and Knowledge Processing).* Springer-Verlag New York, Inc.

Gómez-Pérez, A. (2001). Evaluation of ontologies. *International Journal of intelligent systems.* 16(3)**.** pp. 391-409.

Google. (2011). *Google Lab closure announcement* [Online]. Available from: http://googleblog.blogspot.co.uk/2011/07/more-wood-behind-fewer-arrows.html [Accessed: 11/09/2012].

Grabowicz, P. A., Ramasco, J. J., Moro, E., Pujol, J. M. & Eguiluz, V. M. (2012). Social features of online networks: The strength of intermediary ties in online social media. *PloS one.* 7(1)**.** pp. e29358.

Greene, J. A., Choudhry, N. K., Kilabuk, E. & Shrank, W. H. (2011). Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. *Journal of general internal medicine.* 26(3)**.** pp. 287-292.

Gu, H. H., Elhanan, G., Perl, Y., Hripcsak, G., Cimino, J. J., Xu, J., Chen, Y., Geller, J. & Paul Morrey, C. (2012). A study of terminology auditors' performance for UMLS semantic type assignments. *Journal of biomedical informatics.* 45(6)**.** pp. 1042-1048.

Handcock, M. S. & Gile, K. J. (2011). Comment: On the concept of snowball sampling. *Sociological Methodology.* 41(1)**.** pp. 367-371.

Hanneman, R. A. & Riddle, M. (2005a). Introduction to social network methods. University of California Riverside.

Hanneman, R. A. & Riddle, M. (2005b). Introduction to social network methods. University of California Riverside.

Hardoon, D. R. & Shmueli, G. (2013). *Getting started with business analytics: insightful decision-making.* CRC Press.

Hassan-Montero, Y. & Herrero-Solana, V. (2006). Improving tag-clouds as visual information retrieval interfaces. International Conference on Multidisciplinary Information Sciences and Technologies. Citeseer. pp. 25-28.

Haunschild, P. R. & Miner, A. S. (1997). Modes of interorganizational imitation: The effects of outcome salience and uncertainty. *Administrative science quarterly*. 42(3)**.** pp. 472-500.

Havas, M. (2008). Dirty electricity elevates blood sugar among electrically sensitive diabetics and may explain brittle diabetes. *Electromagnetic biology and medicine.* 27(2)**.** pp. 135-146.

Havre, S., Hetzler, E., Whitney, P. & Nowell, L. (2002). Themeriver: Visualizing thematic changes in large document collections. *Visualization and Computer Graphics, IEEE Transactions on.* 8(1)**.** pp. 9-20.

He, Q. (1999). Knowledge Discovery Through Co-Word Analysis. *Library trends.* 48(1)**.** pp. 133-159.

Hearst, M. A. & Rosner, D. (2008). Tag clouds: Data analysis tool or social signaller? Hawaii International Conference on System Sciences, Proceedings of the 41st Annual. IEEE. pp. 160-170.

Henderson, G. R., Iacobucci, D. & Calder, B. J. (1998). Brand diagnostics: Mapping branding effects using consumer associative networks. *European Journal of Operational Research.* 111(2)**.** pp. 306-327.

Hepp, M. (2010). HyperTwitter: collaborative knowledge engineering via twitter messages. *In:* CIMIANO, P. & PINTO, H. S. (eds.) *Knowledge Engineering and Management by the Masses.* Springer. pp. 451-461.

Herr, P. M., Farquhar, P. H. & Fazio, R. H. (1996). Impact of dominance and relatedness on brand extensions. *Journal of Consumer Psychology.* 5(2)**.** pp. 135-159.

Hogenboom, F., Hogenboom, A., Frasincar, F., Kaymak, U., Van Der Meer, O., Schouten, K. & Vandic, D. (2010). SPEED: A Semantics-Based Pipeline for Economic Event Detection. *Conceptual Modeling–ER 2010.* pp. 452-457.

Hotho, A., Staab, S. & Stumme, G. (2003). Ontologies improve text document clustering. Data Mining, 2003. ICDM 2003. Third IEEE International Conference on. IEEE. pp. 541-544.

Hu, X., Tang, L., Tang, J. & Liu, H. (2013). Exploiting social relations for sentiment analysis in microblogging. *Proceedings of the sixth ACM international conference on Web search and data mining.* Rome, Italy: ACM.

Hu, X., Zhang, X., Lu, C., Park, E. K. & Zhou, X. (2009). Exploiting Wikipedia as external knowledge for document clustering. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining.* Paris, France: ACM.

Huang, A. (2008). Similarity measures for text document clustering. Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008). Christchurch, New Zealand. pp. 49-56.

IHTSDO. (2014). *SNOMED Clinical Terms (SNOMED CT)* [Online]. Available from: http://www.ihtsdo.org/snomed-ct/snomed-ct0/ [Accessed: 22/08/2014].

Iqbal, R., Murad, M. A. A., Mustapha, A. & Sharef, N. M. (2013). An analysis of ontology engineering methodologies: A literature review. *Research Journal of Applied Sciences Engineering and Technology.* 6(16)**.** pp. 2993-3000.

Ivanović, M. & Budimac, Z. (2014). An overview of ontologies and data resources in medical domains. *Expert Systems with Applications.* 41(11)**.** pp. 5158-5166.

Jacob, E. K. (2004). Classification and categorization: a difference that makes a difference.

Jarrar, M. & Meersman, R. (2009). Ontology Engineering - The DOGMA Approach. *In:* THARAM, S. D., ELIZABETH, C., ROBERT, M. & KATIA, S. (eds.) *Advances in Web Semantics I.* Springer-Verlag. pp. 7-34.

John, D. R., Loken, B., Kim, K. & Monga, A. B. (2006). Brand concept maps: a methodology for identifying brand association networks. *Journal of Marketing Research.* 43(4)**.** pp. 549-563.

Kaplan, A. M. & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons.* 53(1)**.** pp. 59-68.

Kim, H., Harris, M. R., Savova, G. & Chute, C. G. (2005). Content coverage of SNOMED-CT toward the ICU nursing flowsheets and the acuity indicators. *Studies in health technology and informatics.* 122(pp. 722-726.

Kim, S.-M., Pantel, P., Chklovski, T. & Pennacchiotti, M. (2006). Automatically assessing review helpfulness. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing.* Sydney, Australia: Association for Computational Linguistics.

Kiryakov, A., Popov, B., Terziev, I., Manov, D. & Ognyanoff, D. (2004). Semantic annotation, indexing, and retrieval. *Web Semantics: Science, Services and Agents on the World Wide Web.* 2(1)**.** pp. 49-79.

Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social networks.* 10(4)**.** pp. 359-381.

Kruse, P., Kummer, C. & Jannack, A. (2015). Empowering Knowledge Transfer in Healthcare: A Framework of Knowledge Transfer Methods. *In: Challenges and Opportunities in Health Care Management.* Springer. pp. 319-328.

Kulkarni, S. & Caragea, D. (2009). Computation of the Semantic Relatedness between Words using Concept Clouds. KDIR. Citeseer. pp. 183-188.

Kumar, P. (2012). Impact of business intelligence systems in Indian telecom industry. *Business Intelligence Journal.* 5(2)**.** pp. 358-366.

Kumaran, G. & Allan, J. (2004). Text classification and named entities for new event detection. Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval. ACM. pp. 297-304.

Kwak, H., Lee, C., Park, H. & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th international conference on World wide web.* Raleigh, North Carolina, USA: ACM.

Lee, D., Cornet, R., Lau, F. & De Keizer, N. (2013). A survey of SNOMED CT implementations. *Journal of biomedical informatics.* 46(1)**.** pp. 87-96.

Li, Y. & Bontcheva, K. (2007). Hierarchical, perceptron-like learning for ontology-based information extraction. Proceedings of the 16th international conference on World Wide Web. ACM. pp. 777-786.

Lin, Y. & Sakamoto, N. (2009). Ontology driven modeling for the knowledge of genetic susceptibility to disease. *Kobe J Med Sci.* 55(3)**.** pp. E53-66.

Liu, H., Wagholikar, K. & Wu, S. T.-I. (2012). Using SNOMED-CT to encode summary level data–a corpus analysis. *AMIA Summits on Translational Science Proceedings.* 2012(pp. 30.

Liu, J., Cao, Y., Lin, C.-Y., Huang, Y. & Zhou, M. (2007). Low-Quality Product Review Detection in Opinion Summarization. *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning EMNLP-CoNLL.*

López-García, P., Boeker, M., Illarramendi, A. & Schulz, S. (2012). Usability-driven pruning of large ontologies: the case of SNOMED CT. *Journal of the American Medical Informatics Association.* 19(e1)**.** pp. e102-e109.

Lu, Y., Tsaparas, P., Ntoulas, A. & Polanyi, L. (2010). Exploiting social context for review quality prediction. *Proceedings of the 19th international conference on World wide web.* Raleigh, North Carolina, USA: ACM.

Ma, X., Bal, J. & Issa, A. (2014). A fast and economic ontology engineering approach towards improving capability matching: Application to an online engineering collaborative platform. *Computers in Industry.*

Majumder, M. & Saha, S. K. (2014). Use of global context for handling noisy names in discussion texts of a homeopathy discussion forum. *Knowledge Management & E-Learning: An International Journal (KM&EL).* 6(1)**.** pp. 18-29.

Mandel, S. E., Davis, B. A. & Secic, M. (2013). Effects of Music Therapy and Music-Assisted Relaxation and Imagery on Health-Related Outcomes in Diabetes Education A Feasibility Study. *The Diabetes Educator.* pp. 0145721713492216.

Manning, C. D., Raghavan, P. & Schütze, H. (2008). *Introduction to information retrieval.* Cambridge university press Cambridge.

McKenzie, J. (2013). *Folksonomy vs. Taxonomy at the Communications Division of Community Services at the City of Edmonton.* MA in Communications and Technology. University of Alberta.

Mendes, P. N., Jakob, M., García-Silva, A. & Bizer, C. (2011). DBpedia spotlight: shedding light on the web of documents. Proceedings of the 7th International Conference on Semantic Systems. ACM. pp. 1-8.

Michelson, M. & Macskassy, S. A. (2010). Discovering users' topics of interest on twitter: a first look. Proceedings of the fourth workshop on Analytics for noisy unstructured text data. ACM. pp. 73-80.

Milne, D. & Witten, I. H. (2008). Learning to link with wikipedia. Proceedings of the 17th ACM conference on Information and knowledge management. ACM. pp. 509-518.

Nachimuthu, S. & Lau, L. M. (2007). Practical issues in using SNOMED CT as a reference terminology. Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems. IOS Press. pp. 640-644.

Nagpaul, P. S. (2003). Exploring a pseudo-regression model of transnational cooperation in science. *Scientometrics.* 56(3)**.** pp. 403-416.

Netzer, O., Feldman, R., Goldenberg, J. & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science.* 31(3)**.** pp. 521-543.

Newman, M. E. (2005). A measure of betweenness centrality based on random walks. *Social networks.* 27(1)**.** pp. 39-54.

Nirenburg, S. & Raskin, V. (2004). *Ontological semantics.* Mit Press.

NLM. (2013). *Unified Medical Language System (UMLS)* [Online]. Available from: http://www.nlm.nih.gov/research/umls/ [Accessed: 22/9/2013].

NLM. (2014). *Medical Subject Headings* [Online]. Available from: http://www.nlm.nih.gov/mesh/ [Accessed: 10/10/2014].

Noy, N. F. & Musen, M. A. (2001). Anchor-PROMPT: Using non-local context for semantic matching. Proceedings of the workshop on ontologies and information sharing at the international joint conference on artificial intelligence (IJCAI). pp. 63-70.

Nyein, S. S. (2011). Mining contents in Web page using cosine similarity. Computer Research and Development (ICCRD), 2011 3rd International Conference on. IEEE. pp. 472-475.

ONS. (2012). *Expenditure on Healthcare in the UK, 2012* [Online]. Available from: http://www.ons.gov.uk/ons/rel/psa/expenditure-on-healthcare-in-the-uk/2012/sty-cost-of-healthcare.html [Accessed: 02/06/2014].

OpenCalais. (2013). *How Does Calais Work?* [Online]. Available from: http://www.opencalais.com/about [Accessed: 22/03/2013].

Özgür, A., Cetin, B. & Bingol, H. (2008). Co-occurrence network of reuters news. *International Journal of Modern Physics C.* 19(05). pp. 689-702.

Pandey, A., Tripathi, P., Pandey, R., Srivatava, R. & Goswami, S. (2011). Alternative therapies useful in the management of diabetes: A systematic review. *Journal of pharmacy & bioallied sciences.* 3(4). pp. 504-512.

Passant, A., Breslin, J. G. & Decker, S. (2010). Rethinking microblogging: open, distributed, semantic. *In:* BENATALLAH, B., CASATI, F., KAPPEL, G. & ROSSI, G. (eds.) *Web Engineering.* Berlin: Springer.

Phan, X.-H., Nguyen, L.-M. & Horiguchi, S. (2008). Learning to classify short and sparse text & web with hidden topics from large-scale data collections. *Proceedings of the 17th international conference on World Wide Web.* Beijing, China: ACM.

Phelan, O., McCarthy, K. & Smyth, B. (2009). Using twitter to recommend real-time topical news. Proceedings of the third ACM conference on Recommender systems. ACM. pp. 385-388.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program.* 14(3). pp. 130-137.

PrincetonUniversity. (2010). *About WordNet* [Online]. Available from: http://wordnet.princeton.edu [Accessed: 22/05/2013].

Ranwez, S., Duthil, B., Sy, M. F., Montmain, J., Augereau, P. & Ranwez, V. (2013). How ontology based information retrieval systems may benefit from lexical text analysis. *In: New Trends of Research in Ontologies and Lexical Resources.* Springer. pp. 209-231.

Rector, A., Rogers, J., Zanstra, P. & Van Der Haring, E. (2003). OpenGALEN: open source medical terminology and tools. AMIA Annual Symposium Proceedings. American Medical Informatics Association. pp. 982.

Rivadeneira, A. W., Gruen, D. M., Muller, M. J. & Millen, D. R. (2007). Getting our head in the clouds: toward evaluation studies of tagclouds. Proceedings of the SIGCHI conference on Human factors in computing systems. ACM. pp. 995-998.

Rizzo, G., Troncy, R., Hellmann, S. & Bruemmer, M. (2012). NERD meets NIF: Lifting NLP Extraction Results to the Linked Data Cloud. *In:* BIZER, C., HEATH, T., BERNERS-LEE, T. & HAUSENBLAS, M., eds. 5th International Workshop on Linked Data on the Web- LDOW2012.

Rogers, J. E. (2004). *Development of a methodology and an ontological schema for medical terminology*. University of Manchester.

Rowe, M. & Stankovic, M. (2012). Aligning tweets with events: Automation via semantics. *Semantic Web.* 3(2)**.** pp. 115-130.

Russell, T. (2006). Cloudalicious: Folksonomy over time. Digital Libraries, 2006. JCDL'06. Proceedings of the 6th ACM/IEEE-CS Joint Conference on. IEEE. pp. 364-364.

Saif, H., He, Y. & Alani, H. (2012). Alleviating data sparsity for twitter sentiment analysis. 2nd Workshop on Making Sense of Microposts (#MSM2012): Big things come in small packages at the 21st International Conference on theWorld Wide Web

(WWW'12). Lyon, France. CEUR Workshop Proceedings (CEUR-WS.org). pp. 2-9.

Saiz, A. & Simonsohn, U. (2008). Downloading wisdom from online crowds. IZA discussion papers.

Salton, G. (1989). *Automatic text processing: the transformation, analysis, and retrieval of information by computer.* Boston, MA, USA. Addison-Wesley Longman Publishing.

Sarasohn-Kahn, J. (2008). *The wisdom of patients: Health care meets online social media.* California HealthCare Foundation Oakland, CA.

Schulz, S., Suntisrivaraporn, B., Baader, F. & Boeker, M. (2009). SNOMED reaching its adolescence: Ontologists' and logicians' health check. *International journal of medical informatics.* 78(pp. S86-S94.

Shahar, Y., Das, A. K., Tu, S. W., Kraemer, F. B. & Musen, M. A. (1994). Knowledge-based temporal abstraction for diabetic monitoring. Proceedings of the Annual Symposium on Computer Application in Medical Care. American Medical Informatics Association. pp. 697.

Shalaby, W., Zadrozny, W. & Gallagher, S. (2014). Knowledge based dimensionality reduction for technical text mining. Big Data (Big Data), 2014 IEEE International Conference on. IEEE. pp. 39-44.

Silva, C. & Ribeiro, B. (2007). On text-based mining with active learning and background knowledge using SVM. *Soft Computing.* 11(6)**.** pp. 519-530.

Singhal, A. (2001). Modern Information Retrieval: A Brief Overview. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering.* 24(4)**.** pp. 35-43.

Sobol-Shikler, T. (2012). Inference of Co-occurring Classes: Multi-class and Multi-label Classification. *In:* OGIELA, M. R. & JAIN, L. C. (eds.) *Computational Intelligence Paradigms in Advanced Pattern Classification.* Springer. pp. 171-197.

Sokolova, M., Japkowicz, N. & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. *In: AI 2006: Advances in Artificial Intelligence.* Springer. pp. 1015-1021.

Specia, L. & Motta, E. (2007). Integrating Folksonomies with the Semantic Web. *The 4th European conference on The Semantic Web: Research and Applications.* Innsbruck, Austria: Springer-Verlag.

Srivastava, A. N. & Sahami, M. (2010). *Text mining: classification, clustering, and applications.* CRC Press.

Staab, S., Studer, R., Schnurr, H.-P. & Sure, Y. (2001). Knowledge processes and ontologies. *IEEE Intelligent systems.* 16(1)**.** pp. 26-34.

Stavrianou, A., Andritsos, P. & Nicoloyannis, N. (2007). Overview and semantic issues of text mining. *ACM Sigmod Record.* 36(3)**.** pp. 23-34.

Stuart, G. & Hulme, C. (2000). The effects of word co-occurance on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* 26(3)**.** pp. 796-802.

Suchanek, F. M., Kasneci, G. & Weikum, G. (2008). Yago: A large ontology from wikipedia and wordnet. *Web Semantics: Science, Services and Agents on the World Wide Web.* 6(3)**.** pp. 203-217.

Swartout, B., Patil, R., Knight, K. & Russ, T. (1996). Toward distributed use of large-scale ontologies. Proceedings of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems. pp. 138-148.

Swartout, B., Ramesh, P., Knight, K. & Russ, T. (eds.) 1997. *Toward Distributed Use of Large-Scale Ontologies.* California: Stanford University.

Teichert, T. A. & Schöntag, K. (2010). Exploring consumer knowledge structures using associative network analysis. *Psychology & Marketing.* 27(4)**.** pp. 369-398.

Turner, V., Gantz, J. F., Reinsel, D. & Minton, S. (2014). The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. International Data Corporation.

Uschold, M. & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The knowledge engineering review.* 11(02)**.** pp. 93-136.

van der Kooij, J., Goossen, W., Goossen-Baremans, A., de Jong-Fintelman, M. & van Beek, L. (2005). Using SNOMED CT codes for coding information in electronic health records for stroke patients. *Studies in health technology and informatics.* 124(pp. 815-823.

Vander Wal, T. (2007). *Folksonomy Coinage and Definition* [Online]. Available from: www.vanderwal.net/folksonomy.html [Accessed: 19 Nov 2013].

Viégas, F. B. & Wattenberg, M. (2008). Timelines tag clouds and the case for vernacular visualization. *interactions.* 15(4)**.** pp. 49-52.

Wang, C., Yu, H. & Ma, K.-L. (2008). Importance-driven time-varying data visualization. *Visualization and Computer Graphics, IEEE Transactions on.* 14(6)**.** pp. 1547-1554.

Wang, X., Tang, L., Gao, H. & Liu, H. (2010). Discovering Overlapping Groups in Social Media. *Proceedings of the 2010 IEEE International Conference on Data Mining.* IEEE Computer Society.

Weinreich, H., Obendorf, H., Herder, E. & Mayer, M. (2008). Not quite the average: An empirical study of Web use. *ACM Transactions on the Web (TWEB).* 2(1)**.** pp. 5.

Welty, C. & Guarino, N. (2001). Supporting ontological analysis of taxonomic relationships. *Data & Knowledge Engineering.* 39(1)**.** pp. 51-74.

WHO (2010). Global status report on noncommunicable diseases 2010. World Health Organization.

WHO. (2013a). *10 facts about diabetes* [Online]. Available from: http://www.who.int/features/factfiles/diabetes/en/ [Accessed: 17/01/2014].

WHO. (2013b). *Noncommunicable diseases - Factsheet* [Online]. Available from: http://www.who.int/mediacentre/factsheets/fs355/en/ [Accessed: 23/04/2014].

Wyner, A., Schneider, J., Atkinson, K. & Bench-Capon, T. J. (2012). Semi-Automated Argumentative Analysis of Online Product Reviews. *COMMA.* 245(pp. 43-50.

Xie, J., Kelley, S. & Szymanski, B. K. (2013). Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Computing Surveys (CSUR).* 45(4)**.** pp. 43-97.

Xu, Z., Fu, Y., Mao, J. & Su, D. (2006). Towards the semantic web: Collaborative tag suggestions. *Collaborative Web Tagging Workshop at WWW2006.* Edinburgh, Scotland.

Yao, Y. (2003). Information-theoretic measures for knowledge discovery and data mining. *In:* KARMESHU (ed.) *Entropy Measures, Maximum Entropy Principle and Emerging Applications.* Springer. pp. 115-136.

Yoo, I., Hu, X. & Song, I.-Y. (2007). Biomedical ontology improves biomedical literature clustering performance: a comparison study. *International journal of bioinformatics research and applications.* 3(3)**.** pp. 414-428.

Zha, Z.-J., Hua, X.-S., Mei, T., Wang, J., Qi, G.-J. & Wang, Z. (2008). Joint multi-label multi-instance learning for image classification. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008).*

Zhang, J. & Zhao, Y. (2013). A user term visualization analysis based on a social question and answer log. *Information Processing & Management.* 49(5)**.** pp. 1019-1048.

Zhang, Y., He, D. & Sang, Y. (2013). Facebook as a platform for health information and communication: a case study of a diabetes group. *Journal of medical systems.* 37(3)**.** pp. 1-12.

Appendices

*Appendixes for this thesis are stored in the CD provided.

Appendix A - Detailed Algorithm for this research's experiments

File location: Appendix\Appendix-A\ Detailed Algorithm.pdf

Appendix B – Data Collected for Diabetes Ontology

This appendix includes two main parts

Appendix B.1 – Degree centrality for Diabetes Ontology

File location: Appendix\Appendix-B\Degree Centrality.pdf

Appendix B.2 – Relationship Calculation for Diabetes Ontology

File location: Appendix\Appendix-B\Relationship Calculation.pdf

Appendix C – Journal Article: A fast and economic ontology engineering approach towards improving capability matching: Application to an online engineering collaborative platform

File location: Appendix\Appendix-C\ Journal Article.pdf

Appendix D – Potential Publications

File location: Appendix\Appendix-D\Potential Publications.pdf

Appendix A – Ontology Corpus Building Algorithm

A.1 Detailed Algorithm for First Round Experiment

- Box 1 feeds the seeding words to OneLook Reverse Dictionary (snowball sampling)
- Box 2 analyses the feedback and forms the results (specify predicted names) and stores in the database for future usage

Similar codes also apply to the second and third round experiments

```
String Newkey1, Newkey2;

Newkey1 = Txb_Keywords1.Text;
Newkey2 = Txb_Keywords2.Text;

Newkey1 = Newkey1.Replace(" ", "+");
Newkey2 = Newkey2.Replace(" ", "+");

string res1 = CheckKeywords(Newkey1, Newkey2);
Lbl_Msg.Text = res1;
Lbl_Msg.Refresh();

if (Lbl_Msg.Text == "Keywords Don't Exist!")
{
    HashSet<string> HT1 = new HashSet<string>();
    HashSet<string> HT2 = new HashSet<string>();
    HashSet<string> Res12 = new HashSet<string>();

    HT1 = GetOutput(Newkey1);
    HT2 = GetOutput(Newkey2);                         1

    Res12 = HT1;
    HT1.Add(Newkey1);
    HT2.Add(Newkey2);

    if (Newkey2 != "")
    {
        Res12.IntersectWith(HT2);
        int pair = Res12.Count;
        storeHashSet(Res12, Newkey1, Newkey2);
    }
    else
    {
        storeHashSet(HT1, Newkey1, Newkey2);
    }
                                                      2
}
```

```
private HashSet<string> GetOutput(string keyword)
        {
            HashSet<string> temp = new HashSet<string>();
            string url = "http://www.onelook.com/?w=*&clue=" + keyword;
            HttpWebRequest HRq = (HttpWebRequest)WebRequest.Create(url);

            HttpWebResponse HRsp = (HttpWebResponse)HRq.GetResponse();

            StreamReader sr = new StreamReader(HRsp.GetResponseStream());

            String HTM = sr.ReadToEnd();

            int pos1 = HTM.IndexOf("1.");

            if (pos1 > 0)
            {
                int Endpos = HTM.IndexOf("</TABLE>", pos1);
                int pos2 = 0;
                int pos3 = 0;


                while (pos1 < Endpos)
                {
                    pos2 = HTM.IndexOf(">", pos1);
                    pos3 = HTM.IndexOf("<", pos2);
                    int length = pos3 - (pos2 + 1);
                    if (HTM.Substring(pos2 + 1, length).Length > 0)
                    {
                        string word = HTM.Substring(pos2 + 1, length);
                        temp.Add(word);
                    }

                    pos1 = HTM.IndexOf("<a", pos3);
                }

                return temp;
            }
            else
            {
                return temp;
            }
        }
```

```
        private void storeHashSet(HashSet<string> HS, string keyword1,
string keyword2)
        {
            string keywords = "";
            foreach (string s in HS)
            {
                keywords += "." + s.Replace(".", ",");
            }

            try
            {
                sqlStr = new StringBuilder();
                sqlStr.Append("insert into ");
                sqlStr.Append(Table1.ToString());
                sqlStr.Append("(keyword1, keyword2, Keywords, predicts)
values ('");
                sqlStr.Append(keyword1.Replace("'", "''") + "','");
                sqlStr.Append(keyword2.Replace("'", "''") + "',");
```

```
                sqlStr.Append(HS.Count.ToString() + ",'");
                sqlStr.Append(keywords.Replace("'", "''") + "')");

                myCommand = new SqlCommand(sqlStr.ToString());
                myCommand.Connection = myConnection;
                myCommand.Connection.Open();
                myCommand.ExecuteNonQuery();
                myCommand.Connection.Close();
            }
            catch (Exception ex)
            {
                MessageBox.Show(this, ex.Message, "Error (Insert into
TableA)", MessageBoxButtons.OK, MessageBoxIcon.Error);
            }

        }
```

Appendix B.1 – Diabetes Ontology Degree Centrality

ID	Keyword	Degree Centrality
1	adult-onset diabetes	794
2	diabetes mellitus	728
3	gestational diabetes	696
4	juvenile diabetes	695
5	type i diabetes	680
6	type ii diabetes	674
7	hyperinsulinism	660
8	adult-onset diabetes mellitus	650
9	autoimmune diabetes	640
10	growth-onset diabetes	635
11	Neonatal Diabetes Mellitus	628
12	insulin-dependent diabetes mellitus	627
13	ketoacidosis-prone diabetes	626
14	ketoacidosis-resistant diabetes	617
15	ketoacidosis-resistant diabetes mellitus	612
16	ketosis-resistant diabetes	610
17	ketosis-prone diabetes	604
18	ketosis-resistant diabetes mellitus	586
19	mature-onset diabetes	585
20	maturity-onset diabetes	580
21	maturity-onset diabetes mellitus	570
22	Niddm	560
23	non-insulin-dependent diabetes	550
24	Tolbutamide	540
25	Dm	530
26	Diabetic	520
27	non-insulin-dependent diabetes mellitus	510
28	Acetonuria	503
29	Banting	477
30	Acetonemia	454
31	ketoaciduria	449
32	Ketonemia	440
33	Insulin	430
34	Orinase	420
35	diabetes insipidus	410
36	antidiabetic drug	400
37	Tolinase	390
38	ketoacidosis prone diabetes	385
39	lente iletin	384
40	lente insulin	380

41	sugar diabetes	376
42	Antidiabetic	370
43	chlorpropamide	366
44	Biguanide	361
45	Ketonuria	355
46	hemochromatosis	353
47	Melituria	351
48	nonketoacidosis prone	351
49	chemical diabetes	350
50	latent diabetes	349
51	Tolazamide	348
52	diabetic diet	347
53	Pancreas	346
54	HbA1C	330
55	Hyperosmolar hyperglycemic nonketotic syndrome	316
56	Basal Rate	310
57	diabetologist	305
58	diabetophobia	290
59	sir frederick grant banting	285
60	nephrogenic diabetes insipidus	280
61	Glucosuria	275
62	Ketosis	270
63	diabetic coma	265
64	Alloxan	257
65	ketone body	250
66	Diabetical	246
67	hyperglycemia	236
68	Diuresis	229
69	diabetic acidosis	225
70	kussmaul's coma	221
71	hypoglycaemic agent	211
72	hypoglycemic agent	207
73	bronzed diabetes	204
74	iron-storage disease	201
75	Juvenile	199
76	sulfonylurea	196
77	iron overload	191
78	Glucophage	190
79	acetoacetic acid	189
80	glucose tolerance test	188
81	somatostatin	176
82	Ketogenesis	175
83	Polydipsia	173
84	brittle diabetes	171

85	necrobiosis lipoidica	168
86	necrobiosis lipoidica diabeticorum	166
87	Lypressin	164
88	hyperglycaemia	161
89	Metformin	157
90	Liable	156
91	Phenformin	151
92	Glycosuria	148
93	ketoacidosis	146
94	carbo loading	138
95	Prediabetes	137
96	carbohydrate loading	133
97	low-calorie diet	128
98	Music	126
99	cushing's disease	125
100	Mellitic	121
101	hypernatremia	120
102	Jambul	119
103	kwashiorkor	113
104	Polyuria	111
105	high-protein diet	108
106	Soft	106
107	Diabeta	105
108	low-fat diet	104
109	Glucagon	100
110	Pap	97
111	insulin reaction	94
112	hypoglycaemia	90
113	beta cell	89
114	Glyburide	88
115	islets of langerhans	87
116	Micronase	86
117	Glipizide	84
118	Pellagra	82
119	Reduce	80
120	insulin shock	79
121	hand-schuller-christian disease	77
122	Pablum	71
123	hypoglycemia	66
124	Glucotrol	65
125	Best	64
126	clear liquid diet	60
127	low-salt diet	59
128	low-sodium diet	58

129	insulin shock therapy	57
130	Sanger	56
131	Lite	55
132	Dietetic	54
133	nicotinic acid	53
134	liquid diet	52
135	salt-free diet	51
136	islet of langerhans	50
137	ketone+body	49
138	Bland	48
139	fad diet	47
140	obesity diet	47
141	reducing diet	47
142	soft diet	47
143	islands of langerhans	47
144	isles of langerhans	46
145	Hodgkin	46
146	Fiber	46
147	bland diet	46
148	high-vitamin diet	45
149	light diet	45
150	humulin	44
151	insulin shock treatment	44
152	macrobiotic	44
153	macleod	43
154	dietetical	42
155	nonfat	41
156	vegetarianism	40
157	allergy diet	39
158	balanced diet	38
159	vitamin-deficiency diet	37
160	xerophagy	37
161	fred sanger	37
162	schuller-christian disease	35
163	diabetes	35
164	low-cal	35
165	train	34
166	fire	34
167	adolescent	34
168	insulin+shock	34
169	dietary	33
170	arginine	33
171	reichstag	32
172	frederick sanger	30

173	cut	30
174	john james rickard macleod	29
175	LCHF	29
176	Fasting plasma glucose test (FPG)	29
177	Gastroparesis	29
178	Lancet	29
179	roughage	28
180	well-fed	26
181	spoon food	26
182	turn	26
183	stripling	25
184	light	25
185	cuttlebone	25
186	regimen	25
187	abstemious	24
188	Retina	24
189	recombinant human insulin	23
190	nurture	23
191	lowcal	22
192	macrobiotic diet	22
193	misdiet	22
194	nocal	22
195	ulcer diet	22
196	augsburg	22
197	bants	22
198	beri	22
199	catamount	22
200	krill	22
201	landtag	22
202	dietetics	21
203	sitology	21
204	john macleod	21
205	strike	20
206	produced	19
207	ketone	19
208	malnutrition	19
209	weaning	19
210	hypoproteinemia	19
211	luminescence	18
212	spare	18
213	xylose	18
214	secretes	17
215	charge	17
216	eat	16

217	nutarian	16
218	victus	16
219	electroconvulsive therapy	16
220	tight control	16
221	electroshock	16
222	metrazol shock	16
223	metrazol shock therapy	16
224	citrulline	16
225	flush	16
226	electroshock therapy	15
227	poor	15
228	take	15
229	teenager	13
230	protein	13
231	ornithine	13
232	flat	13
233	lean	12
234	acetone	12
235	supplement	12
236	vitamin k	12
237	amino+acid	11
238	macrobiotics	11
239	cysteine	11
240	glutamic acid	11
241	glycine	11
242	serine	11
243	thyronine	11
244	peptide	11
245	alanine	11
246	phenylalanine	11
247	proline	11
248	valine	11
249	canavanine	11
250	cystine	11
251	aspartic acid	11
252	histidine	11
253	hydroxyproline	11
254	metrazol shock treatment	11
255	slipper	11
256	break	11
257	methionine	11
258	slim	10
259	food	10
260	diet	10

261	follow	10
262	low	10
263	glucose	10
264	pubescent	10
265	ablactation	9
266	cure	9
267	tyrosine	9
268	dna	9
269	sugar	8
270	gluconeogenesis	8
271	amino acid	8
272	acetone body	8
273	youth	8
274	sweet	8
275	aleuronat	7
276	glycosometer	7
277	age	7
278	coup de main	7
279	surprise attack	7
280	fare	7
281	ascorbic acid	7
282	shock therapy	7
283	glycogen	7
284	cellulose	7
285	monosaccharide	7
286	invert sugar	7
287	lactose	7
288	acetic	7
289	base	7
290	beriberi	7
291	conjugated protein	7
292	twilight	7
293	gegenschein	7
294	blond	7
295	calash	7
296	victoria	7
297	dim	7
298	pale	7
299	pickup	7
300	rockaway	7
301	hertfordshire	7
302	jones' penstemon	7
303	penstemon dolius	7
304	monitor	7

305	lamp	7
306	essential	7
307	endocrine gland	7
308	island of langerhans	7
309	nicotinamide	7
310	pantothen	7
311	spartan	6
312	tofu	6
313	bright	6
314	pitch	6
315	gentle	6
316	flash	6
317	cool	6
318	macerate	6
319	nyctalopia	6
320	bitartrate	6
321	saponify	6
322	avoid	6
323	protestant	6
324	dieted	6
325	positive	6
326	form	6
327	rear	5
328	force	5
329	sweetbread	5
330	blitz	5
331	adolescence	5
332	access	5
333	down	5
334	foot	5
335	avitaminosis	5
336	ants	5
337	sea	5
338	starch	5
339	preventive	5
340	shock+therapy	5
341	choline	5
342	fall	5
343	spar	5
344	riboflavin	5
345	asparagine	5
346	fat	5
347	galvanic	5
348	glow	5

349	shade	4
350	shine	4
351	beam	4
352	chiaroscuro	4
353	counterglow	4
354	dawn	4
355	firelight	4
356	flare	4
357	gleam	4
358	halo	4
359	illuminate	4
360	lens	4
361	ray	4
362	refraction	4
363	shaft	4
364	starlight	4
365	sun	4
366	electrocute	4
367	air spring	4
368	electrify	4
369	sonic boom	4
370	air cushion	4
371	revolt	4
372	shockproof	4
373	bump	4
374	galvanize	4
375	stagger	4
376	jar	4
377	dashpot	4
378	slender	4
379	blank	4
380	bude light	4
381	flux	4
382	deficiency disease	4
383	owe	4
384	sago	4
385	smooth	4
386	mild	4
387	dumb	4
388	recover	4
389	ting	4
390	folacin	4
391	hypocalcaemia	4
392	hypocalcemia	4

393	rickets	4
394	starve	4
395	tryptophan	4
396	tryptophane	4
397	first class	4
398	pride	4
399	calcium	4
400	iron	4
401	leucine	4
402	rda	4
403	mush	4
404	sucrose	4
405	seminose	4
406	beet sugar	4
407	thiamine	4
408	idiot light	4
409	Oral diabetes medications	4
410	candlelight	4
411	evil	4
412	dark	4
413	faint	4
414	Insulin pump	4
415	scurvy	4
416	simple	4
417	narrow	4
418	mellow	4
419	yeast	4
420	emetic	4
421	radiotherapy	4
422	principal	4
423	cathartic	4
424	dose	4
425	acetic acid	4
426	sour	4
427	pathogen	4
428	stearic acid	4
429	subacid	4
430	succinic acid	4
431	tartrate	4
432	urate	4
433	Stevia	4
434	acid-forming	4
435	manganic acid	4
436	margaric acid	4

437	mucic acid	4
438	nitric acid	4
439	oil of vitriol	4
440	oleic acid	4
441	oxyacid	4
442	paba	4
443	phenol	4
444	phosphorous acid	4
445	phthalic acid	4
446	racemic acid	4
447	salicylate	4
448	hydrochloric acid	4
449	ethanedioic acid	4
450	fatty acid	4
451	formic acid	4
452	gallic acid	4
453	glyceric acid	4
454	heptadecanoic acid	4
455	cerotic acid	4
456	chloroacetic acid	4
457	chromate	4
458	chromic acid	4
459	cyanamide	4
460	vinegar	4
461	acetyl	4
462	acidophilic	4
463	acidulous	4
464	arsenate	4
465	phosphate	4
466	vitriol	4
467	selenic acid	4
468	picric acid	4
469	rna	4
470	aqua fortis	4
471	etch	4
472	nitrate	4
473	superphosphate	4
474	acetify	4
475	acidify	4
476	acidity	4
477	boric acid	4
478	butyric acid	4
479	citrate	4
480	cyanic acid	4

481	linolenic acid	4
482	fumaric acid	4
483	oxalic acid	4
484	saccharic acid	4
485	sulphate	4
486	senate	4
487	house	4
488	sound	4
489	propanone	4
490	glycose	4
491	corpse	4
492	short	4
493	black	4
494	starvation acidosis	4
495	dieting	4
496	legislature	4
497	pip	4
498	plague	4
499	lupus	4
500	influenza	4
501	get	4
502	aspirin	4
503	brown	4
504	abortifacient	4
505	depressant	4
506	anodyne	4
507	hydrogen cyanide	4
508	ives	4
509	evans	4
510	parry	4
511	churchill	4
512	make	4
513	character	4
514	catch	4
515	counterattack	4
516	diversion	4
517	incursion	4
518	banzai attack	4
519	bombardment	4
520	penetration	4
521	invasion	4
522	girl	4
523	ketotic	4
524	check	4

525	prepuberty	4
526	run	4
527	classic	4
528	flower	4
529	mass	4
530	adult	4
531	absolute	4
532	condition	3
533	jejune	3
534	prime	3
535	grow	3
536	page	3
537	block	3
538	come	3
539	man	3
540	drag	3
541	order	3
542	anticonvulsant	3
543	complex	3
544	communicate	3
545	antiphlogistic	3
546	beneficial	3
547	arbitrary	3
548	carbohydrate	3
549	right	3
550	planet	3
551	star	3
552	rust	3
553	chlorosis	3
554	ferment	3
555	propranolol	3
556	cross	3
557	christian science	3
558	back	3
559	drain	3
560	fasting	3
561	easy	3
562	sodium	3
563	dip	3
564	table	3
565	course	3
566	weak	3
567	cane sugar	3
568	deficiency+disease	3

569	sucrate	3
570	porter	3
571	dog	3
572	place	3
573	feed	3
574	napier	3
575	reducing	3
576	pantothenic acid	3
577	tail	3
578	gluten	3
579	geophagy	3
580	shadow	3
581	horrify	3
582	white	3
583	aurora	2
584	blind	2
585	dazzle	2
586	exposure	2
587	floodlight	2
588	fluorescence	2
589	gig	2
590	glare	2
591	glimmer	2
592	glitter	2
593	kite	2
594	lantern	2
595	luminous	2
596	moon	2
597	moonlight	2
598	opaque	2
599	photic	2
600	radiant	2
601	reflect	2
602	sidelight	2
603	spotlight	2
604	twinkle	2
605	wherry	2
606	cays	2
607	continental	2
608	go	2
609	clerk	2
610	carry	2
611	accompany	2
612	coach	2

613	groom	2
614	educate	2
615	cultivate	2
616	discipline	2
617	drill	2
618	exercise	2
619	bubble	2
620	rinse	2
621	chyle	2
622	ether	2
623	salt pork	2
624	bolt	2
625	freight	2
626	freight liner	2
627	save	2
628	conducive	2
629	result	2
630	laser	2
631	dry	2
632	vary	2
633	guard	2
634	board	2
635	blandly	2
636	blandness	2
637	insipid	2
638	suave	2
639	blanding	2
640	flavorless	2
641	flavourless	2
642	savorless	2
643	vapid	2
644	vanilla	2
645	arbutus unedo	2
646	bromide	2
647	cottage+cheese	2
648	drastic	2
649	faair-spoken	2
650	favonian	2
651	impregnably	2
652	irish strawberry	2
653	muenster	2
654	obtundent	2
655	pop	2
656	pop music	2

657	strawberry tree	2
658	tame	2
659	tasteless	2
660	til seed	2
661	unctuous	2
662	watery	2
663	conference	2
664	gland	2
665	rearward	2
666	hebrides	2
667	electron	2
668	distributive shock	2
669	obstructive shock	2
670	slam	2
671	fright	2
672	shake up	2
673	cardiogenic shock	2
674	earthquake	2
675	electric	2
676	hormephobia	2
677	hypovolemic shock	2
678	torpillage	2
679	metrazol	2
680	pentamethylenetetrazol	2
681	pentylenetetrazol	2
682	sensitive	2
683	defibrillation	2
684	serum albumin	2
685	electroshock+therapy	2
686	galvanism	2
687	shock treatment	2
688	merry	2
689	cruiser	2
690	badminton	2
691	cot	2
692	strobe lighting	2
693	top+boot	2
694	skyrocket	2
695	hydroplane	2
696	spallation	2
697	horse latitudes	2
698	synchrotron radiation	2
699	polymer	2
700	incandescence	2

701	incandescent	2
702	lighthouse	2
703	top-boots	2
704	white-hot	2
705	mao jacket	2
706	spider phaeton	2
707	horse+latitudes	2
708	process plate	2
709	relax	2
710	fusion bomb	2
711	h-bomb	2
712	hydrogen bomb	2
713	thermonuclear bomb	2
714	phase-contrast microscope	2
715	krypton	2
716	express rifle	2
717	aluminum	2
718	joy	2
719	kerr cell	2
720	els	2
721	zip	2
722	codon	2
723	dopa	2
724	isoleucine	2
725	threonine	2
726	glutamine	2
727	essential amino acid	2
728	lysine	2
729	gaba	2
730	gamma aminobutyric acid	2
731	glutaminic acid	2
732	iodoamino acid	2
733	sarcosine	2
734	transfer rna	2
735	taurine	2
736	inessential amino acid	2
737	degenerate	2
738	trna	2
739	genetic code	2
740	glutathione	2
741	dihydroxyphenylalanine	2
742	triiodothyronine	2
743	peptide bond	2
744	creatine	2

745	creatin	2
746	degeneracy	2
747	ethionine	2
748	glutamic+acid	2
749	pentapeptide	2
750	isoelectric point	2
751	tyramine	2
752	structural+gene	2
753	translate	2
754	thyroxine	2
755	anticodon	2
756	lysine intolerance	2
757	nonsense	2
758	lysinemia	2
759	triplet code	2
760	acceptor rna	2
761	soluble rna	2
762	ribosome	2
763	polypeptide	2
764	aminoalkanoic acid	2
765	messenger rna	2
766	diamine	2
767	peptone	2
768	sequence	2
769	deamination	2
770	deaminization	2
771	aminopherase	2
772	aminotransferase	2
773	carbamino	2
774	transaminase	2
775	glutamate	2
776	linoleic acid	2
777	aminoplast	2
778	deaminate	2
779	linolic acid	2
780	transaminate	2
781	transamination	2
782	supply	2
783	code	2
784	hypovitaminosis	2
785	rachitis	2
786	tocopherol	2
787	night blindness	2
788	moon blindness	2

789	hypothrombinemia	2
790	biotin	2
791	cobalamin	2
792	cyanocobalamin	2
793	antipernicious anemia factor	2
794	niacin	2
795	nicotinic+acid	2
796	clean	2
797	clear	2
798	happy	2
799	manograph	2
800	antibody	2
801	genus chlorella	2
802	rocket	2
803	blue	2
804	determination	2
805	silver	2
806	simplicity	2
807	recess	2
808	fibre	2
809	reduction	2
810	wesleyan	2
811	car	2
812	convention	2
813	wilkes	2
814	wesley	2
815	watson	2
816	marshall	2
817	byng	2
818	spot	2
819	salt	2
820	shoot	2
821	mark	2
822	work	2
823	father	2
824	sulfur	2
825	whig	2
826	eliot	2
827	stuart	2
828	canning	2
829	hit	2
830	record	2
831	tiptop	2
832	paranymph	2

833	elite	2
834	pick	2
835	second-best	2
836	chart	2
837	rollo	2
838	snow	2
839	primary	2
840	scoop	2
841	level	2
842	aqua regia	2
843	osteopetrosis	2
844	herpes	2
845	albers-schonberg disease	2
846	hyperadrenalism	2
847	better	2
848	cream	2
849	optimum	2
850	aristocracy	2
851	triticin	2
852	candy	2
853	praline	2
854	saccharate	2
855	saccharification	2
856	saccharify	2
857	saccharose	2
858	sinistrin	2
859	kelp	2
860	nutraceutical	2
861	vitamin pill	2
862	liver	2
863	mannitol	2
864	creeps	2
865	nonessential	2
866	mash	2
867	flummery	2
868	adipocere	2
869	fluff	2
870	fluffy	2
871	poplar	2
872	flannel	2
873	levulosan	2
874	maple sugar	2
875	mucilage	2
876	ribose	2

877	saccharide	2
878	carbohydrates	2
879	#NAME?	2
880	polysaccharide	2
881	deoxyribose	2
882	glycoprotein	2
883	jaggary	2
884	jaggery	2
885	jagghery	2
886	monosaccharose	2
887	oligosaccharide	2
888	polyose	2
889	agar	2
890	amylum	2
891	antigen	2
892	arabin	2
893	chemosynthesis	2
894	citric acid	2
895	comfort food	2
896	cortef	2
897	corticosteroid	2
898	corticosterone	2
899	cortisol	2
900	dark+reaction	2
901	dextrin	2
902	fermentation	2
903	galactin	2
904	gelose	2
905	glyceraldehyde	2
906	glycolipid	2
907	hydrate	2
908	hydrocortisone	2
909	hydrocortone	2
910	lactic+acid	2
911	lignin	2
912	mucoid	2
913	pectose	2
914	phosphofructokinase	2
915	refined sugar	2
916	whisper	2
917	gondola	2
918	cocoa powder	2
919	tennis	2
920	elves	2

921	kart	2
922	burley	2
923	cold light	2
924	hansom cab	2
925	tobogganing	2
926	dimmer	2
927	pickup truck	2
928	proxima	2
929	chemiluminescence	2
930	leucocratic	2
931	airglow	2
932	toboggan	2
933	phototaxy	2
934	reichsrath	2
935	virga	2
936	oats	2
937	sunscald	2
938	coffee	2
939	chloroplast	2
940	bowl	2
941	thin	2
942	twist	2
943	plate	2
944	sup	2
945	indulge	2
946	eats	2
947	hash	2
948	repeat	2
949	spider	2
950	help	2
951	kernite	2
952	downy	2
953	zephyr	2
954	challis	2
955	candle	2
956	lithium	2
957	quash	2
958	breakfast	2
959	cannibalize	2
960	beat	2
961	overload	2
962	power loading	2
963	stevedore	2
964	lighterage	2

965	save-all	2
966	trainload	2
967	autoloading	2
968	burthen	2
969	cascabel	2
970	dead load	2
971	fraughtage	2
972	fraughting	2
973	hopper	2
974	live load	2
975	millstone	2
976	oneration	2
977	span loading	2
978	stevedorage	2
979	superload	2
980	dispatch	2
981	crop	2
982	avert	2
983	derail	2
984	fence	2
985	nouvelle cuisine	2
986	spark	2
987	jump	2
988	shave	2
989	lopsided	2
990	burn	2
991	slenderize	2
992	trim	2
993	contract	2
994	decrease	2
995	reducer	2
996	pass	2
997	berenice's hair	2
998	carus	2
999	envelope	2
1000	hepatic coma	2
1001	semicomatose	2
1002	sign	2
1003	comatose	2
1004	comate	2
1005	narcoma	2
1006	but	2
1007	christen	2
1008	over	2

1009	crown wart	2
1010	heaven	2
1011	nematode	2
1012	crown gall	2
1013	molluscum	2
1014	new thought	2
1015	agapeistic	2
1016	doctor	2
1017	sacrament	2
1018	receive	2
1019	kiss of peace	2
1020	pesthouse	2
1021	erythema	2
1022	albigensianism	2
1023	catharism	2
1024	inri	2
1025	mennonite	2
1026	faith cure	2
1027	kiss+of+peace	2
1028	symbolism	2
1029	martyrdom	2
1030	coenurus	2
1031	glycaemia	2
1032	maltase	2
1033	amylase	2
1034	corn sugar	2
1035	adrenaline	2
1036	gluconic	2
1037	phosphorylase	2
1038	commission	2
1039	dextrose	2
1040	inversion	2
1041	glucoside	2
1042	grape sugar	2
1043	invertase	2
1044	dextroglucose	2
1045	blood glucose	2
1046	blood sugar	2
1047	hyperalimentation	2
1048	total parenteral nutrition	2
1049	tpn	2
1050	comae	2
1051	comatic	2
1052	comatoseness	2

1053	convulsions	2
1054	torpor	2
1055	narcotic	2
1056	lead poisoning	2
1057	comet	2
1058	heatstroke	2
1059	sleep	2
1060	nucleus	2
1061	eclampsia	2
1062	braxy	2
1063	coma berenices	2
1064	comose	2
1065	karen ann quinlan	2
1066	revive	2
1067	robin cook	2
1068	schmidt telescope	2
1069	semicoma	2
1070	emulsifier	2
1071	engine	2
1072	estate agent	2
1073	federal official	2
1074	general agent	2
1075	house agent	2
1076	infection	2
1077	intravenous anesthetic	2
1078	land agent	2
1079	local anaesthetic	2
1080	local anesthetic	2
1081	mutagen	2
1082	resolvent	2
1083	spinal anaesthetic	2
1084	spinal anesthetic	2
1085	steward	2
1086	surfactant	2
1087	syndic	2
1088	t-man	2
1089	teratogen	2
1090	thinner	2
1091	topical anaesthetic	2
1092	topical anesthetic	2
1093	active	2
1094	alcahest	2
1095	alkahest	2
1096	alkalizer	2

1097	antacid	2
1098	antidote	2
1099	bacteriostat	2
1100	bailee	2
1101	bailiff	2
1102	blanching agent	2
1103	bleaching agent	2
1104	local	2
1105	disinfectant	2
1106	realtor	2
1107	bleach	2
1108	attorney	2
1109	detergent	2
1110	procurator	2
1111	lenient	2
1112	cerebrospinal meningitis	2
1113	maltose	2
1114	envelop	2
1115	mydriasis	2
1116	glycogenesis	2
1117	hexose	2
1118	amylose	2
1119	galactose	2
1120	sucrase	2
1121	lactase	2
1122	cirrhosis	2
1123	agent provocateur	2
1124	chemical agent	2
1125	clorox	2
1126	coolant	2
1127	decoagulant	2
1128	deus ex machina	2
1129	dissolvent	2
1130	dissolver	2
1131	expectorant	2
1132	fungicide	2
1133	gastric antacid	2
1134	soap	2
1135	fed	2
1136	g-man	2
1137	spiritual being	2
1138	supernatural being	2
1139	bleaching powder	2
1140	cause	2

1141	cautery	2
1142	chloride of lime	2
1143	federal	2
1144	germicide	2
1145	vasoconstrictor	2
1146	whitener	2
1147	anticoagulant	2
1148	antifungal	2
1149	antimycotic	2
1150	business agent	2
1151	comprador	2
1152	desiccant	2
1153	diluent	2
1154	dilutant	2
1155	puddle	2
1156	greensickness	2
1157	pig	2
1158	bug	2
1159	brand	2
1160	hold	2
1161	burden	2
1162	overburden	2
1163	read	2
1164	terminal	2
1165	penicillamine	2
1166	store	2
1167	hook	2
1168	acidulate	2
1169	person	2
1170	#NAME?	2
1171	water	2
1172	flesh	2
1173	church	2
1174	constituency	2
1175	sphinx	2
1176	ball	2
1177	cast iron	2
1178	wood sugar	2
1179	spleen	2
1180	ary	2
1181	aschaffenburg	2
1182	botanical medicine	2
1183	athletic training	2
1184	touch	2

1185	cult	2
1186	leprosy	2
1187	sequela	2
1188	aspergillosis	2
1189	clap	2
1190	barrie	2
1191	subject	2
1192	labile	2
1193	suspicious	2
1194	mutable	2
1195	peccable	2
1196	apt	2
1197	likely	2
1198	syphilis	2
1199	emmenagogue	2
1200	prescription	2
1201	painkiller	2
1202	placebo	2
1203	astringent	2
1204	pharmaceutical	2
1205	physic	2
1206	antiviral	2
1207	pain pill	2
1208	fleming	2
1209	murray	2
1210	hall	2
1211	morgan	2
1212	wallace	2
1213	ross	2
1214	thomson	2
1215	grey	2
1216	fluorite	2
1217	fluorspar	2
1218	pancreatic	2
1219	islets	2
1220	bismuth	2
1221	weld	2
1222	dewar	2
1223	ashton	2
1224	hopkins	2
1225	pollock	2
1226	whitaker	2
1227	ondine	2
1228	herschel	2

1229	north	2
1230	wilkins	2
1231	barbarossa	2
1232	frederick the great	2
1233	frederiksberg	2
1234	handel	2
1235	ketose	2
1236	ketosteroid	2
1237	keto	2
1238	anthraquinone	2
1239	testosterone	2
1240	ketohexose	2
1241	flavin	2
1242	butyrone	2
1243	laurone	2
1244	margarone	2
1245	myristone	2
1246	oenanthone	2
1247	palmitone	2
1248	propione	2
1249	suberone	2
1250	valerone	2
1251	xanthone	2
1252	flavanone	2
1253	ketonic	2
1254	benzoin	2
1255	stearone	2
1256	benzophenone	2
1257	dimethyl ketone	2
1258	hexone	2
1259	thienone	2
1260	acetophenone	2
1261	methyl isobutyl ketone	2
1262	androsterone	2
1263	camphor	2
1264	pregnenolone	2
1265	benzoquinone	2
1266	butanone	2
1267	ketoxime	2
1268	methyl ethyl ketone	2
1269	cinnamone	2
1270	oleone	2
1271	quinone	2
1272	ketone group	2

1273	chassis	2
1274	state	2
1275	transit	2
1276	passage	2
1277	somatic	2
1278	anatomy	2
1279	frame	2
1280	orb	2
1281	college	2
1282	corpus	2
1283	figure	2
1284	government	2
1285	member	2
1286	shape	2
1287	system	2
1288	adult body	2
1289	carcass	2
1290	incorporate	2
1291	physique	2
1292	skin	2
1293	soma	2
1294	temperature	2
1295	carbonyl	2
1296	hemiacetal	2
1297	clay	2
1298	corselet	2
1299	impression	2
1300	child	2
1301	kid	2
1302	fry	2
1303	nestling	2
1304	youngster	2
1305	younker	2
1306	nipper	2
1307	shaver	2
1308	small fry	2
1309	juvenility	2
1310	tike	2
1311	tyke	2
1312	minor	2
1313	tiddler	2
1314	young person	2
1315	delinquent	2
1316	juvenile delinquency	2

1317	ingenue	2
1318	juvenile delinquent	2
1319	delinquency	2
1320	hebesphalmology	2
1321	juvenile court	2
1322	apple box	2
1323	coordinate	2
1324	generation	2
1325	stick	2
1326	perfect	2
1327	throw	2
1328	chip	2
1329	logotype	2
1330	division	2
1331	diiodide	2
1332	pithecanthropus	2
1333	face	2
1334	lead	2
1335	space	2
1336	woman	2
1337	imago	2
1338	bull	2
1339	stag	2
1340	tenor	2
1341	bombing	2
1342	horse	2
1343	eve	2
1344	madam	2
1345	warmonger	2
1346	banzai charge	2
1347	teen	2
1348	boy	2
1349	homology	2
1350	edition	2
1351	autumn	2
1352	haggard	2
1353	indehiscent	2
1354	maturate	2
1355	prime of life	2
1356	ripeness	2
1357	rising	2
1358	bud	2
1359	seed	2
1360	call	2

1361	pancreatic juice	2
1362	pancreatin	2
1363	cf	2
1364	banteng	2
1365	bos banteng	2
1366	tsine	2
1367	bantingism	2
1368	diastase	2
1369	amylopsin	2
1370	grown-up	2
1371	ripe	2
1372	womanhood	2
1373	coccus	2
1374	emerging	2
1375	independent	2
1376	contingent	2
1377	anaclitic	2
1378	adjective	2
1379	varistor	2
1380	white man's burden	2
1381	free	2
1382	conditional	2
1383	un-	2
1384	conventional	2
1385	contingency	2
1386	relative	2
1387	growth	2
1388	mature	2
1389	implicit	2
1390	secondary	2
1391	satellite	2
1392	precarious	2
1393	gentile	1
1394	extraneous	1
1395	light-headed	1
1396	raise	1
1397	discount	1
1398	cystic fibrosis	1
1399	hoyle	1
1400	lento	1
1401	condensation	1
1402	shank	1
1403	rave	1
1404	stud	1

1405	blight	1
1406	stand	1
1407	show	1
1408	attack	1
1409	leverage	1
1410	letter	1
1411	countermove	1
1412	sow	1
1413	scleroderma	1
1414	baby	1
1415	weed	1
1416	rise	1
1417	child's body	1
1418	breed	1
1419	bond	1
1420	strain	1
1421	type a	1
1422	council	1
1423	arsenic	1
1424	fluor	1
1425	range	1
1426	gb	1
1427	conversion	1
1428	maurice	1
1429	barkley	1
1430	acton	1
1431	peel	1
1432	dope	1
1433	banks	1
1434	purgative	1
1435	medication	1
1436	amphetamine	1
1437	power	1
1438	antihistamine	1
1439	acid	1
1440	susceptible	1
1441	indulgent	1
1442	dispose	1
1443	antispasmodic	1
1444	remedy	1
1445	cocaine	1
1446	head	1
1447	use	1
1448	resistance	1

1449	apocalypse	1
1450	anemia	1
1451	celiac	1
1452	repress	1
1453	develop	1
1454	cementite	1
1455	be	1
1456	matzoh	1
1457	friable	1
1458	osteoporosis	1
1459	answer	1
1460	open	1
1461	equity	1
1462	grate	1
1463	stuff	1
1464	zinc deficiency	1
1465	bloom	1
1466	sponge	1
1467	blow	1
1468	release	1
1469	choke	1
1470	chalybeate	1
1471	necrobiotic	1
1472	pack	1
1473	drum	1
1474	boot	1
1475	hard	1
1476	virus	1
1477	set	1
1478	closet	1
1479	go-between	1
1480	fructose	1
1481	abetalipoproteinemia	1
1482	try	1
1483	barth	1
1484	fehling's+solution	1
1485	wear	1
1486	kindly	1
1487	unconscious	1
1488	relapse	1
1489	cell	1
1490	surgery	1
1491	treat	1
1492	grind	1

1493	deplete	1
1494	slough	1
1495	dismount	1
1496	stop	1
1497	lift	1
1498	drab	1
1499	moderate	1
1500	deflect	1
1501	drop	1
1502	liquidate	1
1503	less	1
1504	bite	1
1505	wing loading	1
1506	do	1
1507	abstain	1
1508	get off	1
1509	deficiency	1
1510	skim	1
1511	fudge	1
1512	fill	1
1513	victual	1
1514	dilute	1
1515	mess	1
1516	delicate	1
1517	duck	1
1518	cormorant	1
1519	spoon-fed	1
1520	thread	1
1521	bruise	1
1522	domestic	1
1523	plastic	1
1524	deep	1
1525	reflex	1
1526	allergic reaction	1
1527	equilibrium	1
1528	pulp	1
1529	bran	1
1530	dietetically	1
1531	wide	1
1532	canker	1
1533	specific	1
1534	lamb	1
1535	green	1
1536	marble bones disease	1

1537	barbituric acid	1
1538	basic	1
1539	good	1
1540	moore	1
1541	coke	1
1542	weber	1
1543	give	1
1544	hack	1
1545	kind	1
1546	match	1
1547	burke	1
1548	buchanan	1
1549	hunt	1
1550	st	1
1551	succeed	1
1552	charles i	1
1553	mary ii	1
1554	halifax	1
1555	lion	1
1556	methodists	1
1557	hydrolysis	1
1558	fusion	1
1559	scorbutus	1
1560	pilot	1
1561	wanton	1
1562	auxiliary	1
1563	flab	1
1564	support	1
1565	baptist	1
1566	draw	1
1567	fatness	1
1568	lesion	1
1569	wire	1
1570	vitamin b12	1
1571	gastric	1
1572	#NAME?	1
1573	resolution	1
1574	febrifuge	1
1575	supersensitive	1
1576	red	1
1577	radiation	1
1578	buffer	1
1579	transformation	1
1580	aran	1

1581	isle	1
1582	issue	1
1583	life	1
1584	tonneau	1
1585	accidental	1
1586	luther	1
1587	silk	1
1588	bill	1
1589	devil	1
1590	ration	1
1591	bundesrath	1
1592	floater	1
1593	court	1
1594	scant	1
1595	heel	1
1596	petrolatum	1
1597	practice	1
1598	purge	1
1599	cotton	1
1600	cushion	1
1601	chancroid	1
1602	poultice	1
1603	natural	1
1604	benefit	1
1605	serve	1
1606	healthful	1
1607	genetic engineering	1
1608	liner train	1
1609	tailgate	1
1610	lighter	1
1611	wet	1
1612	obese	1
1613	gluten-free diet	1
1614	articulate	1
1615	string	1
1616	manage	1
1617	tack	1
1618	watch	1
1619	ascetic	1
1620	europe	1
1621	ait	1
1622	window	1

Appendix B.2 – Diabetes Ontology Relationship Calculation

ID	baseTerm	targetTerm	baseTermCount	targetTermCount	ratio
1	diabetes mellitus	adult-onset diabetes	39	753	0.051793
2	gestational diabetes	adult-onset diabetes	65	753	0.086321
3	juvenile diabetes	adult-onset diabetes	39	753	0.051793
4	type i diabetes	adult-onset diabetes	38	753	0.050465
5	type ii diabetes	adult-onset diabetes	27	753	0.035857
6	adult-onset diabetes mellitus	adult-onset diabetes	39	753	0.051793
7	autoimmune diabetes	adult-onset diabetes	39	753	0.051793
8	growth-onset diabetes	adult-onset diabetes	38	753	0.050465
9	insulin-dependent diabetes mellitus	adult-onset diabetes	39	753	0.051793
10	ketoacidosis-prone diabetes	adult-onset diabetes	39	753	0.051793
11	ketoacidosis-resistant diabetes	adult-onset diabetes	39	753	0.051793
12	ketoacidosis-resistant diabetes mellitus	adult-onset diabetes	39	753	0.051793
13	ketosis-resistant diabetes	adult-onset diabetes	39	753	0.051793
14	ketosis-prone diabetes	adult-onset diabetes	39	753	0.051793
15	ketosis-resistant diabetes mellitus	adult-onset diabetes	39	753	0.051793
16	mature-onset diabetes	adult-onset diabetes	39	753	0.051793
17	maturity-onset diabetes	adult-onset diabetes	38	753	0.050465
18	maturity-onset diabetes mellitus	adult-onset diabetes	39	753	0.051793

19	non-insulin-dependent diabetes	adult-onset diabetes	38	753	0.050465
20	non-insulin-dependent diabetes mellitus	adult-onset diabetes	39	753	0.051793
21	insulin	adult-onset diabetes	38	753	0.050465
22	diabetes insipidus	adult-onset diabetes	39	753	0.051793
23	ketoacidosis prone diabetes	adult-onset diabetes	39	753	0.051793
24	lente insulin	adult-onset diabetes	38	753	0.050465
25	sugar diabetes	adult-onset diabetes	39	753	0.051793
26	chemical diabetes	adult-onset diabetes	39	753	0.051793
27	latent diabetes	adult-onset diabetes	39	753	0.051793
28	nephrogenic diabetes insipidus	adult-onset diabetes	39	753	0.051793
29	hypoglycemic agent	adult-onset diabetes	37	753	0.049137
30	bronzed diabetes	adult-onset diabetes	38	753	0.050465
31	insulin reaction	adult-onset diabetes	37	753	0.049137
32	insulin shock	adult-onset diabetes	73	753	0.096946
33	insulin shock therapy	adult-onset diabetes	37	753	0.049137
34	obesity diet	adult-onset diabetes	37	753	0.049137
35	soft diet	adult-onset diabetes	34	753	0.045153
36	bland diet	adult-onset diabetes	37	753	0.049137
37	diabetes	adult-onset diabetes	38	753	0.050465
38	recombinant human insulin	adult-onset diabetes	37	753	0.049137
39	adult-onset diabetes	diabetes mellitus	36	689	0.05225
40	gestational diabetes	diabetes mellitus	63	689	0.091437
41	juvenile diabetes	diabetes mellitus	37	689	0.053701
42	type i diabetes	diabetes mellitus	37	689	0.053701
43	type ii diabetes	diabetes mellitus	36	689	0.05225
44	adult-onset diabetes mellitus	diabetes mellitus	36	689	0.05225
45	autoimmune diabetes	diabetes mellitus	37	689	0.053701
46	growth-onset diabetes	diabetes mellitus	37	689	0.053701

47	insulin-dependent diabetes mellitus	diabetes mellitus	37	689	0.053701
48	ketoacidosis-prone diabetes	diabetes mellitus	37	689	0.053701
49	ketoacidosis-resistant diabetes	diabetes mellitus	37	689	0.053701
50	ketoacidosis-resistant diabetes mellitus	diabetes mellitus	37	689	0.053701
51	ketosis-resistant diabetes	diabetes mellitus	37	689	0.053701
52	ketosis-prone diabetes	diabetes mellitus	37	689	0.053701
53	ketosis-resistant diabetes mellitus	diabetes mellitus	37	689	0.053701
54	mature-onset diabetes	diabetes mellitus	36	689	0.05225
55	maturity-onset diabetes	diabetes mellitus	37	689	0.053701
56	maturity-onset diabetes mellitus	diabetes mellitus	36	689	0.05225
57	non-insulin-dependent diabetes	diabetes mellitus	37	689	0.053701
58	non-insulin-dependent diabetes mellitus	diabetes mellitus	37	689	0.053701
59	insulin	diabetes mellitus	37	689	0.053701
60	diabetes insipidus	diabetes mellitus	37	689	0.053701
61	ketoacidosis prone diabetes	diabetes mellitus	37	689	0.053701
62	lente insulin	diabetes mellitus	37	689	0.053701
63	sugar diabetes	diabetes mellitus	38	689	0.055152
64	chemical diabetes	diabetes mellitus	37	689	0.053701
65	latent diabetes	diabetes mellitus	37	689	0.053701
66	nephrogenic diabetes insipidus	diabetes mellitus	37	689	0.053701

67	bronzed diabetes	diabetes mellitus	36	689	0.05225
68	carbohydrate loading	diabetes mellitus	1	689	0.001451
69	polyuria	diabetes mellitus	34	689	0.049347
70	insulin reaction	diabetes mellitus	36	689	0.05225
71	insulin shock	diabetes mellitus	71	689	0.103048
72	insulin shock therapy	diabetes mellitus	31	689	0.044993
73	insulin shock treatment	diabetes mellitus	34	689	0.049347
74	diabetes	diabetes mellitus	36	689	0.05225
75	recombinant human insulin	diabetes mellitus	37	689	0.053701
76	carbohydrate	diabetes mellitus	3	689	0.004354
77	adult-onset diabetes	gestational diabetes	35	630	0.055556
78	diabetes mellitus	gestational diabetes	35	630	0.055556
79	juvenile diabetes	gestational diabetes	35	630	0.055556
80	type i diabetes	gestational diabetes	35	630	0.055556
81	type ii diabetes	gestational diabetes	35	630	0.055556
82	adult-onset diabetes mellitus	gestational diabetes	35	630	0.055556
83	autoimmune diabetes	gestational diabetes	35	630	0.055556
84	growth-onset diabetes	gestational diabetes	35	630	0.055556
85	insulin-dependent diabetes mellitus	gestational diabetes	35	630	0.055556
86	ketoacidosis-prone diabetes	gestational diabetes	35	630	0.055556
87	ketoacidosis-resistant diabetes	gestational diabetes	35	630	0.055556
88	ketoacidosis-resistant diabetes mellitus	gestational diabetes	35	630	0.055556
89	ketosis-resistant diabetes	gestational diabetes	35	630	0.055556
90	ketosis-prone diabetes	gestational diabetes	35	630	0.055556
91	ketosis-resistant	gestational diabetes	35	630	0.055556

	diabetes mellitus				
92	mature-onset diabetes	gestational diabetes	35	630	0.055556
93	maturity-onset diabetes	gestational diabetes	35	630	0.055556
94	maturity-onset diabetes mellitus	gestational diabetes	35	630	0.055556
95	non-insulin-dependent diabetes	gestational diabetes	35	630	0.055556
96	non-insulin-dependent diabetes mellitus	gestational diabetes	35	630	0.055556
97	insulin	gestational diabetes	35	630	0.055556
98	diabetes insipidus	gestational diabetes	35	630	0.055556
99	ketoacidosis prone diabetes	gestational diabetes	35	630	0.055556
100	lente insulin	gestational diabetes	35	630	0.055556
101	sugar diabetes	gestational diabetes	35	630	0.055556
102	chemical diabetes	gestational diabetes	35	630	0.055556
103	latent diabetes	gestational diabetes	35	630	0.055556
104	nephrogenic diabetes insipidus	gestational diabetes	35	630	0.055556
105	bronzed diabetes	gestational diabetes	35	630	0.055556
106	ketoacidosis	gestational diabetes	35	630	0.055556
107	polyuria	gestational diabetes	35	630	0.055556
108	insulin reaction	gestational diabetes	35	630	0.055556
109	insulin shock	gestational diabetes	69	630	0.109524
110	diabetes	gestational diabetes	35	630	0.055556
111	recombinant human insulin	gestational diabetes	35	630	0.055556
112	adult-onset diabetes	juvenile diabetes	36	658	0.054711
113	diabetes mellitus	juvenile diabetes	36	658	0.054711
114	gestational diabetes	juvenile diabetes	63	658	0.095745
115	type i diabetes	juvenile diabetes	36	658	0.054711
116	type ii diabetes	juvenile diabetes	36	658	0.054711
117	adult-onset diabetes mellitus	juvenile diabetes	36	658	0.054711

118	autoimmune diabetes	juvenile diabetes	36	658	0.054711
119	growth-onset diabetes	juvenile diabetes	36	658	0.054711
120	insulin-dependent diabetes mellitus	juvenile diabetes	36	658	0.054711
121	ketoacidosis-prone diabetes	juvenile diabetes	36	658	0.054711
122	ketoacidosis-resistant diabetes	juvenile diabetes	36	658	0.054711
123	ketoacidosis-resistant diabetes mellitus	juvenile diabetes	36	658	0.054711
124	ketosis-resistant diabetes	juvenile diabetes	36	658	0.054711
125	ketosis-prone diabetes	juvenile diabetes	36	658	0.054711
126	ketosis-resistant diabetes mellitus	juvenile diabetes	36	658	0.054711
127	mature-onset diabetes	juvenile diabetes	36	658	0.054711
128	maturity-onset diabetes	juvenile diabetes	36	658	0.054711
129	maturity-onset diabetes mellitus	juvenile diabetes	36	658	0.054711
130	non-insulin-dependent diabetes	juvenile diabetes	36	658	0.054711
131	non-insulin-dependent diabetes mellitus	juvenile diabetes	36	658	0.054711
132	insulin	juvenile diabetes	36	658	0.054711
133	diabetes insipidus	juvenile diabetes	36	658	0.054711
134	ketoacidosis prone diabetes	juvenile diabetes	36	658	0.054711
135	lente insulin	juvenile diabetes	36	658	0.054711
136	sugar diabetes	juvenile diabetes	36	658	0.054711
137	chemical diabetes	juvenile diabetes	36	658	0.054711
138	latent diabetes	juvenile diabetes	36	658	0.054711

139	nephrogenic diabetes insipidus	juvenile diabetes	36	658	0.054711
140	bronzed diabetes	juvenile diabetes	35	658	0.053191
141	ketoacidosis	juvenile diabetes	36	658	0.054711
142	polyuria	juvenile diabetes	35	658	0.053191
143	insulin reaction	juvenile diabetes	35	658	0.053191
144	insulin shock	juvenile diabetes	69	658	0.104863
145	diabetes	juvenile diabetes	35	658	0.053191
146	recombinant human insulin	juvenile diabetes	35	658	0.053191
147	adult-onset diabetes	type i diabetes	36	644	0.055901
148	diabetes mellitus	type i diabetes	36	644	0.055901
149	gestational diabetes	type i diabetes	63	644	0.097826
150	juvenile diabetes	type i diabetes	36	644	0.055901
151	type ii diabetes	type i diabetes	34	644	0.052795
152	adult-onset diabetes mellitus	type i diabetes	36	644	0.055901
153	autoimmune diabetes	type i diabetes	36	644	0.055901
154	growth-onset diabetes	type i diabetes	36	644	0.055901
155	insulin-dependent diabetes mellitus	type i diabetes	36	644	0.055901
156	ketoacidosis-prone diabetes	type i diabetes	36	644	0.055901
157	ketoacidosis-resistant diabetes	type i diabetes	36	644	0.055901
158	ketoacidosis-resistant diabetes mellitus	type i diabetes	36	644	0.055901
159	ketosis-resistant diabetes	type i diabetes	36	644	0.055901
160	ketosis-prone diabetes	type i diabetes	36	644	0.055901
161	ketosis-resistant diabetes mellitus	type i diabetes	36	644	0.055901

162	mature-onset diabetes	type i diabetes	36	644	0.055901
163	maturity-onset diabetes	type i diabetes	36	644	0.055901
164	maturity-onset diabetes mellitus	type i diabetes	36	644	0.055901
165	non-insulin-dependent diabetes	type i diabetes	36	644	0.055901
166	non-insulin-dependent diabetes mellitus	type i diabetes	36	644	0.055901
167	insulin	type i diabetes	36	644	0.055901
168	diabetes insipidus	type i diabetes	36	644	0.055901
169	ketoacidosis prone diabetes	type i diabetes	36	644	0.055901
170	lente insulin	type i diabetes	36	644	0.055901
171	sugar diabetes	type i diabetes	36	644	0.055901
172	chemical diabetes	type i diabetes	36	644	0.055901
173	latent diabetes	type i diabetes	36	644	0.055901
174	nephrogenic diabetes insipidus	type i diabetes	36	644	0.055901
175	bronzed diabetes	type i diabetes	34	644	0.052795
176	ketoacidosis	type i diabetes	28	644	0.043478
177	polyuria	type i diabetes	28	644	0.043478
178	insulin reaction	type i diabetes	33	644	0.051242
179	insulin shock	type i diabetes	65	644	0.100932
180	diabetes	type i diabetes	33	644	0.051242
181	recombinant human insulin	type i diabetes	33	644	0.051242
182	adult-onset diabetes	type ii diabetes	38	673	0.056464
183	diabetes mellitus	type ii diabetes	37	673	0.054978
184	gestational diabetes	type ii diabetes	64	673	0.095097
185	juvenile diabetes	type ii diabetes	38	673	0.056464
186	type i diabetes	type ii diabetes	6	673	0.008915
187	adult-onset diabetes mellitus	type ii diabetes	38	673	0.056464

188	autoimmune diabetes	type ii diabetes	37	673	0.054978
189	growth-onset diabetes	type ii diabetes	33	673	0.049034
190	insulin-dependent diabetes mellitus	type ii diabetes	37	673	0.054978
191	ketoacidosis-prone diabetes	type ii diabetes	37	673	0.054978
192	ketoacidosis-resistant diabetes	type ii diabetes	37	673	0.054978
193	ketoacidosis-resistant diabetes mellitus	type ii diabetes	37	673	0.054978
194	ketosis-resistant diabetes	type ii diabetes	37	673	0.054978
195	ketosis-prone diabetes	type ii diabetes	37	673	0.054978
196	ketosis-resistant diabetes mellitus	type ii diabetes	37	673	0.054978
197	mature-onset diabetes	type ii diabetes	38	673	0.056464
198	maturity-onset diabetes	type ii diabetes	35	673	0.052006
199	maturity-onset diabetes mellitus	type ii diabetes	37	673	0.054978
200	non-insulin-dependent diabetes	type ii diabetes	37	673	0.054978
201	non-insulin-dependent diabetes mellitus	type ii diabetes	37	673	0.054978
202	insulin	type ii diabetes	31	673	0.046062
203	diabetes insipidus	type ii diabetes	37	673	0.054978
204	ketoacidosis prone diabetes	type ii diabetes	37	673	0.054978
205	lente insulin	type ii diabetes	31	673	0.046062
206	sugar diabetes	type ii diabetes	37	673	0.054978
207	chemical diabetes	type ii diabetes	37	673	0.054978
208	latent diabetes	type ii diabetes	37	673	0.054978

209	nephrogenic diabetes insipidus	type ii diabetes	37	673	0.054978
210	hypoglycemic agent	type ii diabetes	27	673	0.040119
211	bronzed diabetes	type ii diabetes	35	673	0.052006
212	insulin reaction	type ii diabetes	34	673	0.05052
213	insulin shock	type ii diabetes	71	673	0.105498
214	insulin shock therapy	type ii diabetes	32	673	0.047548
215	obesity diet	type ii diabetes	30	673	0.044577
216	bland diet	type ii diabetes	31	673	0.046062
217	diabetes	type ii diabetes	35	673	0.052006
218	recombinant human insulin	type ii diabetes	32	673	0.047548
219	adult-onset diabetes	hyperinsulinism	35	641	0.054602
220	diabetes mellitus	hyperinsulinism	35	641	0.054602
221	gestational diabetes	hyperinsulinism	61	641	0.095164
222	juvenile diabetes	hyperinsulinism	36	641	0.056162
223	adult-onset diabetes mellitus	hyperinsulinism	34	641	0.053042
224	autoimmune diabetes	hyperinsulinism	36	641	0.056162
225	growth-onset diabetes	hyperinsulinism	34	641	0.053042
226	insulin-dependent diabetes mellitus	hyperinsulinism	36	641	0.056162
227	ketoacidosis-prone diabetes	hyperinsulinism	36	641	0.056162
228	ketoacidosis-resistant diabetes	hyperinsulinism	36	641	0.056162
229	ketoacidosis-resistant diabetes mellitus	hyperinsulinism	35	641	0.054602
230	ketosis-resistant diabetes	hyperinsulinism	36	641	0.056162
231	ketosis-prone diabetes	hyperinsulinism	36	641	0.056162

232	ketosis-resistant diabetes mellitus	hyperinsulinism	35	641	0.054602
233	mature-onset diabetes	hyperinsulinism	35	641	0.054602
234	maturity-onset diabetes	hyperinsulinism	35	641	0.054602
235	maturity-onset diabetes mellitus	hyperinsulinism	34	641	0.053042
236	non-insulin-dependent diabetes	hyperinsulinism	35	641	0.054602
237	non-insulin-dependent diabetes mellitus	hyperinsulinism	34	641	0.053042
238	insulin	hyperinsulinism	37	641	0.057722
239	diabetes insipidus	hyperinsulinism	36	641	0.056162
240	ketoacidosis prone diabetes	hyperinsulinism	36	641	0.056162
241	lente insulin	hyperinsulinism	37	641	0.057722
242	sugar diabetes	hyperinsulinism	36	641	0.056162
243	chemical diabetes	hyperinsulinism	36	641	0.056162
244	latent diabetes	hyperinsulinism	34	641	0.053042
245	pancreas	hyperinsulinism	32	641	0.049922
246	nephrogenic diabetes insipidus	hyperinsulinism	36	641	0.056162
247	bronzed diabetes	hyperinsulinism	35	641	0.054602
248	insulin reaction	hyperinsulinism	36	641	0.056162
249	hypoglycaemia	hyperinsulinism	4	641	0.00624
250	insulin shock	hyperinsulinism	69	641	0.107644
251	hypoglycemia	hyperinsulinism	19	641	0.029641
252	insulin shock treatment	hyperinsulinism	35	641	0.054602
253	diabetes	hyperinsulinism	35	641	0.054602
254	recombinant human insulin	hyperinsulinism	34	641	0.053042
255	adult-onset diabetes	adult-onset diabetes mellitus	35	594	0.058923
256	diabetes mellitus	adult-onset diabetes mellitus	35	594	0.058923
257	gestational diabetes	adult-onset diabetes mellitus	55	594	0.092593

258	juvenile diabetes	adult-onset diabetes mellitus	34	594	0.057239
259	autoimmune diabetes	adult-onset diabetes mellitus	34	594	0.057239
260	insulin-dependent diabetes mellitus	adult-onset diabetes mellitus	35	594	0.058923
261	ketoacidosis-prone diabetes	adult-onset diabetes mellitus	34	594	0.057239
262	ketoacidosis-resistant diabetes	adult-onset diabetes mellitus	34	594	0.057239
263	ketoacidosis-resistant diabetes mellitus	adult-onset diabetes mellitus	35	594	0.058923
264	ketosis-resistant diabetes	adult-onset diabetes mellitus	34	594	0.057239
265	ketosis-prone diabetes	adult-onset diabetes mellitus	34	594	0.057239
266	ketosis-resistant diabetes mellitus	adult-onset diabetes mellitus	35	594	0.058923
267	mature-onset diabetes	adult-onset diabetes mellitus	35	594	0.058923
268	maturity-onset diabetes mellitus	adult-onset diabetes mellitus	35	594	0.058923
269	non-insulin-dependent diabetes	adult-onset diabetes mellitus	35	594	0.058923
270	non-insulin-dependent diabetes mellitus	adult-onset diabetes mellitus	35	594	0.058923
271	insulin	adult-onset diabetes mellitus	35	594	0.058923
272	diabetes insipidus	adult-onset diabetes mellitus	34	594	0.057239
273	ketoacidosis prone diabetes	adult-onset diabetes mellitus	34	594	0.057239
274	lente insulin	adult-onset diabetes mellitus	35	594	0.058923
275	sugar diabetes	adult-onset diabetes mellitus	34	594	0.057239
276	chemical diabetes	adult-onset diabetes mellitus	34	594	0.057239

277	latent diabetes	adult-onset diabetes mellitus	11	594	0.018519
278	nephrogenic diabetes insipidus	adult-onset diabetes mellitus	34	594	0.057239
279	hypoglycemic agent	adult-onset diabetes mellitus	33	594	0.055556
280	bronzed diabetes	adult-onset diabetes mellitus	33	594	0.055556
281	insulin reaction	adult-onset diabetes mellitus	33	594	0.055556
282	insulin shock	adult-onset diabetes mellitus	65	594	0.109428
283	insulin shock therapy	adult-onset diabetes mellitus	33	594	0.055556
284	obesity diet	adult-onset diabetes mellitus	33	594	0.055556
285	soft diet	adult-onset diabetes mellitus	33	594	0.055556
286	bland diet	adult-onset diabetes mellitus	33	594	0.055556
287	diabetes	adult-onset diabetes mellitus	33	594	0.055556
288	recombinant human insulin	adult-onset diabetes mellitus	33	594	0.055556
289	adult-onset diabetes	autoimmune diabetes	34	587	0.057922
290	diabetes mellitus	autoimmune diabetes	34	587	0.057922
291	gestational diabetes	autoimmune diabetes	59	587	0.100511
292	juvenile diabetes	autoimmune diabetes	34	587	0.057922
293	adult-onset diabetes mellitus	autoimmune diabetes	34	587	0.057922
294	growth-onset diabetes	autoimmune diabetes	34	587	0.057922
295	insulin-dependent diabetes mellitus	autoimmune diabetes	34	587	0.057922
296	ketoacidosis-prone diabetes	autoimmune diabetes	34	587	0.057922
297	ketoacidosis-resistant diabetes	autoimmune diabetes	34	587	0.057922
298	ketoacidosis-resistant diabetes mellitus	autoimmune diabetes	34	587	0.057922

299	ketosis-resistant diabetes	autoimmune diabetes	34	587	0.057922
300	ketosis-prone diabetes	autoimmune diabetes	34	587	0.057922
301	ketosis-resistant diabetes mellitus	autoimmune diabetes	34	587	0.057922
302	mature-onset diabetes	autoimmune diabetes	34	587	0.057922
303	maturity-onset diabetes	autoimmune diabetes	34	587	0.057922
304	maturity-onset diabetes mellitus	autoimmune diabetes	34	587	0.057922
305	non-insulin-dependent diabetes	autoimmune diabetes	34	587	0.057922
306	non-insulin-dependent diabetes mellitus	autoimmune diabetes	34	587	0.057922
307	insulin	autoimmune diabetes	34	587	0.057922
308	diabetes insipidus	autoimmune diabetes	34	587	0.057922
309	ketoacidosis prone diabetes	autoimmune diabetes	34	587	0.057922
310	lente insulin	autoimmune diabetes	34	587	0.057922
311	sugar diabetes	autoimmune diabetes	34	587	0.057922
312	chemical diabetes	autoimmune diabetes	34	587	0.057922
313	latent diabetes	autoimmune diabetes	34	587	0.057922
314	nephrogenic diabetes insipidus	autoimmune diabetes	34	587	0.057922
315	bronzed diabetes	autoimmune diabetes	33	587	0.056218
316	ketoacidosis	autoimmune diabetes	34	587	0.057922
317	polyuria	autoimmune diabetes	33	587	0.056218
318	insulin reaction	autoimmune diabetes	33	587	0.056218
319	insulin shock	autoimmune diabetes	65	587	0.110733
320	diabetes	autoimmune diabetes	33	587	0.056218
321	recombinant human insulin	autoimmune diabetes	33	587	0.056218
322	adult-onset diabetes	growth-onset diabetes	34	587	0.057922
323	diabetes mellitus	growth-onset diabetes	34	587	0.057922

324	gestational diabetes	growth-onset diabetes	59	587	0.100511
325	juvenile diabetes	growth-onset diabetes	34	587	0.057922
326	adult-onset diabetes mellitus	growth-onset diabetes	34	587	0.057922
327	autoimmune diabetes	growth-onset diabetes	34	587	0.057922
328	insulin-dependent diabetes mellitus	growth-onset diabetes	34	587	0.057922
329	ketoacidosis-prone diabetes	growth-onset diabetes	34	587	0.057922
330	ketoacidosis-resistant diabetes	growth-onset diabetes	34	587	0.057922
331	ketoacidosis-resistant diabetes mellitus	growth-onset diabetes	34	587	0.057922
332	ketosis-resistant diabetes	growth-onset diabetes	34	587	0.057922
333	ketosis-prone diabetes	growth-onset diabetes	34	587	0.057922
334	ketosis-resistant diabetes mellitus	growth-onset diabetes	34	587	0.057922
335	mature-onset diabetes	growth-onset diabetes	34	587	0.057922
336	maturity-onset diabetes	growth-onset diabetes	34	587	0.057922
337	maturity-onset diabetes mellitus	growth-onset diabetes	34	587	0.057922
338	non-insulin-dependent diabetes	growth-onset diabetes	34	587	0.057922
339	non-insulin-dependent diabetes mellitus	growth-onset diabetes	34	587	0.057922
340	insulin	growth-onset diabetes	34	587	0.057922
341	diabetes insipidus	growth-onset diabetes	34	587	0.057922
342	ketoacidosis prone diabetes	growth-onset diabetes	34	587	0.057922

343	lente insulin	growth-onset diabetes	34	587	0.057922
344	sugar diabetes	growth-onset diabetes	34	587	0.057922
345	chemical diabetes	growth-onset diabetes	34	587	0.057922
346	latent diabetes	growth-onset diabetes	34	587	0.057922
347	nephrogenic diabetes insipidus	growth-onset diabetes	34	587	0.057922
348	bronzed diabetes	growth-onset diabetes	33	587	0.056218
349	ketoacidosis	growth-onset diabetes	34	587	0.057922
350	polyuria	growth-onset diabetes	33	587	0.056218
351	insulin reaction	growth-onset diabetes	33	587	0.056218
352	insulin shock	growth-onset diabetes	65	587	0.110733
353	diabetes	growth-onset diabetes	33	587	0.056218
354	recombinant human insulin	growth-onset diabetes	33	587	0.056218
355	adult-onset diabetes	Neonatal Diabetes Mellitus	35	622	0.05627
356	diabetes mellitus	Neonatal Diabetes Mellitus	35	622	0.05627
357	gestational diabetes	Neonatal Diabetes Mellitus	61	622	0.098071
358	juvenile diabetes	Neonatal Diabetes Mellitus	35	622	0.05627
359	adult-onset diabetes mellitus	Neonatal Diabetes Mellitus	35	622	0.05627
360	autoimmune diabetes	Neonatal Diabetes Mellitus	35	622	0.05627
361	growth-onset diabetes	Neonatal Diabetes Mellitus	35	622	0.05627
362	insulin-dependent diabetes mellitus	Neonatal Diabetes Mellitus	35	622	0.05627
363	ketoacidosis-prone diabetes	Neonatal Diabetes Mellitus	35	622	0.05627
364	ketoacidosis-resistant diabetes	Neonatal Diabetes Mellitus	35	622	0.05627
365	ketoacidosis-resistant diabetes mellitus	Neonatal Diabetes Mellitus	35	622	0.05627
366	ketosis-resistant diabetes	Neonatal Diabetes Mellitus	35	622	0.05627
367	ketosis-prone diabetes	Neonatal Diabetes Mellitus	35	622	0.05627

368	ketosis-resistant diabetes mellitus	Neonatal Diabetes Mellitus	35	622	0.05627
369	mature-onset diabetes	Neonatal Diabetes Mellitus	35	622	0.05627
370	maturity-onset diabetes	Neonatal Diabetes Mellitus	35	622	0.05627
371	maturity-onset diabetes mellitus	Neonatal Diabetes Mellitus	35	622	0.05627
372	non-insulin-dependent diabetes	Neonatal Diabetes Mellitus	35	622	0.05627
373	non-insulin-dependent diabetes mellitus	Neonatal Diabetes Mellitus	35	622	0.05627
374	insulin	Neonatal Diabetes Mellitus	35	622	0.05627
375	diabetes insipidus	Neonatal Diabetes Mellitus	35	622	0.05627
376	ketoacidosis prone diabetes	Neonatal Diabetes Mellitus	35	622	0.05627
377	lente insulin	Neonatal Diabetes Mellitus	35	622	0.05627
378	sugar diabetes	Neonatal Diabetes Mellitus	35	622	0.05627
379	chemical diabetes	Neonatal Diabetes Mellitus	35	622	0.05627
380	latent diabetes	Neonatal Diabetes Mellitus	35	622	0.05627
381	nephrogenic diabetes insipidus	Neonatal Diabetes Mellitus	35	622	0.05627
382	bronzed diabetes	Neonatal Diabetes Mellitus	34	622	0.054662
383	ketoacidosis	Neonatal Diabetes Mellitus	35	622	0.05627
384	polyuria	Neonatal Diabetes Mellitus	34	622	0.054662
385	insulin reaction	Neonatal Diabetes Mellitus	34	622	0.054662
386	insulin shock	Neonatal Diabetes Mellitus	67	622	0.107717
387	diabetes	Neonatal Diabetes Mellitus	34	622	0.054662
388	recombinant human insulin	Neonatal Diabetes Mellitus	34	622	0.054662
389	adult-onset diabetes	insulin-dependent diabetes mellitus	34	587	0.057922

390	diabetes mellitus	insulin-dependent diabetes mellitus	34	587	0.057922
391	gestational diabetes	insulin-dependent diabetes mellitus	59	587	0.100511
392	juvenile diabetes	insulin-dependent diabetes mellitus	34	587	0.057922
393	adult-onset diabetes mellitus	insulin-dependent diabetes mellitus	34	587	0.057922
394	autoimmune diabetes	insulin-dependent diabetes mellitus	34	587	0.057922
395	growth-onset diabetes	insulin-dependent diabetes mellitus	34	587	0.057922
396	ketoacidosis-prone diabetes	insulin-dependent diabetes mellitus	34	587	0.057922
397	ketoacidosis-resistant diabetes	insulin-dependent diabetes mellitus	34	587	0.057922
398	ketoacidosis-resistant diabetes mellitus	insulin-dependent diabetes mellitus	34	587	0.057922
399	ketosis-resistant diabetes	insulin-dependent diabetes mellitus	34	587	0.057922
400	ketosis-prone diabetes	insulin-dependent diabetes mellitus	34	587	0.057922
401	ketosis-resistant diabetes mellitus	insulin-dependent diabetes mellitus	34	587	0.057922
402	mature-onset diabetes	insulin-dependent diabetes mellitus	34	587	0.057922
403	maturity-onset diabetes	insulin-dependent diabetes mellitus	34	587	0.057922
404	maturity-onset diabetes mellitus	insulin-dependent diabetes mellitus	34	587	0.057922
405	non-insulin-dependent diabetes	insulin-dependent diabetes mellitus	34	587	0.057922
406	non-insulin-dependent diabetes mellitus	insulin-dependent diabetes mellitus	34	587	0.057922
407	insulin	insulin-dependent diabetes mellitus	34	587	0.057922
408	diabetes insipidus	insulin-dependent diabetes mellitus	34	587	0.057922
409	ketoacidosis prone diabetes	insulin-dependent diabetes mellitus	34	587	0.057922

410	lente insulin	insulin-dependent diabetes mellitus	34	587	0.057922
411	sugar diabetes	insulin-dependent diabetes mellitus	34	587	0.057922
412	chemical diabetes	insulin-dependent diabetes mellitus	34	587	0.057922
413	latent diabetes	insulin-dependent diabetes mellitus	34	587	0.057922
414	nephrogenic diabetes insipidus	insulin-dependent diabetes mellitus	34	587	0.057922
415	bronzed diabetes	insulin-dependent diabetes mellitus	33	587	0.056218
416	ketoacidosis	insulin-dependent diabetes mellitus	34	587	0.057922
417	polyuria	insulin-dependent diabetes mellitus	33	587	0.056218
418	insulin reaction	insulin-dependent diabetes mellitus	33	587	0.056218
419	insulin shock	insulin-dependent diabetes mellitus	65	587	0.110733
420	diabetes	insulin-dependent diabetes mellitus	33	587	0.056218
421	recombinant human insulin	insulin-dependent diabetes mellitus	33	587	0.056218
422	adult-onset diabetes	ketoacidosis-prone diabetes	34	587	0.057922
423	diabetes mellitus	ketoacidosis-prone diabetes	34	587	0.057922
424	gestational diabetes	ketoacidosis-prone diabetes	59	587	0.100511
425	juvenile diabetes	ketoacidosis-prone diabetes	34	587	0.057922
426	adult-onset diabetes mellitus	ketoacidosis-prone diabetes	34	587	0.057922
427	autoimmune diabetes	ketoacidosis-prone diabetes	34	587	0.057922
428	growth-onset diabetes	ketoacidosis-prone diabetes	34	587	0.057922
429	insulin-dependent diabetes mellitus	ketoacidosis-prone diabetes	34	587	0.057922
430	ketoacidosis-resistant diabetes	ketoacidosis-prone diabetes	34	587	0.057922
431	ketoacidosis-resistant diabetes mellitus	ketoacidosis-prone diabetes	34	587	0.057922

432	ketosis-resistant diabetes	ketoacidosis-prone diabetes	34	587	0.057922
433	ketosis-prone diabetes	ketoacidosis-prone diabetes	34	587	0.057922
434	ketosis-resistant diabetes mellitus	ketoacidosis-prone diabetes	34	587	0.057922
435	mature-onset diabetes	ketoacidosis-prone diabetes	34	587	0.057922
436	maturity-onset diabetes	ketoacidosis-prone diabetes	34	587	0.057922
437	maturity-onset diabetes mellitus	ketoacidosis-prone diabetes	34	587	0.057922
438	non-insulin-dependent diabetes	ketoacidosis-prone diabetes	34	587	0.057922
439	non-insulin-dependent diabetes mellitus	ketoacidosis-prone diabetes	34	587	0.057922
440	insulin	ketoacidosis-prone diabetes	34	587	0.057922
441	diabetes insipidus	ketoacidosis-prone diabetes	34	587	0.057922
442	ketoacidosis prone diabetes	ketoacidosis-prone diabetes	34	587	0.057922
443	lente insulin	ketoacidosis-prone diabetes	34	587	0.057922
444	sugar diabetes	ketoacidosis-prone diabetes	34	587	0.057922
445	chemical diabetes	ketoacidosis-prone diabetes	34	587	0.057922
446	latent diabetes	ketoacidosis-prone diabetes	34	587	0.057922
447	nephrogenic diabetes insipidus	ketoacidosis-prone diabetes	34	587	0.057922
448	bronzed diabetes	ketoacidosis-prone diabetes	33	587	0.056218
449	ketoacidosis	ketoacidosis-prone diabetes	34	587	0.057922
450	polyuria	ketoacidosis-prone diabetes	33	587	0.056218
451	insulin reaction	ketoacidosis-prone diabetes	33	587	0.056218
452	insulin shock	ketoacidosis-prone diabetes	65	587	0.110733

453	diabetes	ketoacidosis-prone diabetes	33	587	0.056218
454	recombinant human insulin	ketoacidosis-prone diabetes	33	587	0.056218
455	adult-onset diabetes	ketoacidosis-resistant diabetes	34	583	0.058319
456	diabetes mellitus	ketoacidosis-resistant diabetes	34	583	0.058319
457	gestational diabetes	ketoacidosis-resistant diabetes	55	583	0.09434
458	juvenile diabetes	ketoacidosis-resistant diabetes	34	583	0.058319
459	adult-onset diabetes mellitus	ketoacidosis-resistant diabetes	34	583	0.058319
460	autoimmune diabetes	ketoacidosis-resistant diabetes	34	583	0.058319
461	insulin-dependent diabetes mellitus	ketoacidosis-resistant diabetes	34	583	0.058319
462	ketoacidosis-prone diabetes	ketoacidosis-resistant diabetes	34	583	0.058319
463	ketoacidosis-resistant diabetes mellitus	ketoacidosis-resistant diabetes	34	583	0.058319
464	ketosis-resistant diabetes	ketoacidosis-resistant diabetes	34	583	0.058319
465	ketosis-prone diabetes	ketoacidosis-resistant diabetes	34	583	0.058319
466	ketosis-resistant diabetes mellitus	ketoacidosis-resistant diabetes	34	583	0.058319
467	mature-onset diabetes	ketoacidosis-resistant diabetes	34	583	0.058319
468	maturity-onset diabetes mellitus	ketoacidosis-resistant diabetes	34	583	0.058319
469	non-insulin-dependent diabetes	ketoacidosis-resistant diabetes	34	583	0.058319
470	non-insulin-dependent diabetes mellitus	ketoacidosis-resistant diabetes	34	583	0.058319
471	insulin	ketoacidosis-resistant diabetes	34	583	0.058319

472	diabetes insipidus	ketoacidosis-resistant diabetes	34	583	0.058319
473	ketoacidosis prone diabetes	ketoacidosis-resistant diabetes	34	583	0.058319
474	lente insulin	ketoacidosis-resistant diabetes	34	583	0.058319
475	sugar diabetes	ketoacidosis-resistant diabetes	34	583	0.058319
476	chemical diabetes	ketoacidosis-resistant diabetes	34	583	0.058319
477	nephrogenic diabetes insipidus	ketoacidosis-resistant diabetes	34	583	0.058319
478	hypoglycemic agent	ketoacidosis-resistant diabetes	33	583	0.056604
479	bronzed diabetes	ketoacidosis-resistant diabetes	33	583	0.056604
480	insulin reaction	ketoacidosis-resistant diabetes	33	583	0.056604
481	insulin shock	ketoacidosis-resistant diabetes	65	583	0.111492
482	insulin shock therapy	ketoacidosis-resistant diabetes	33	583	0.056604
483	obesity diet	ketoacidosis-resistant diabetes	33	583	0.056604
484	soft diet	ketoacidosis-resistant diabetes	33	583	0.056604
485	bland diet	ketoacidosis-resistant diabetes	33	583	0.056604
486	diabetes	ketoacidosis-resistant diabetes	33	583	0.056604
487	recombinant human insulin	ketoacidosis-resistant diabetes	33	583	0.056604
488	adult-onset diabetes	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
489	diabetes mellitus	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
490	gestational diabetes	ketoacidosis-resistant diabetes mellitus	55	583	0.09434
491	juvenile diabetes	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
492	adult-onset diabetes mellitus	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
493	autoimmune diabetes	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
494	insulin-dependent diabetes mellitus	ketoacidosis-resistant diabetes mellitus	34	583	0.058319

495	ketoacidosis-prone diabetes	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
496	ketoacidosis-resistant diabetes	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
497	ketosis-resistant diabetes	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
498	ketosis-prone diabetes	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
499	ketosis-resistant diabetes mellitus	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
500	mature-onset diabetes	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
501	maturity-onset diabetes mellitus	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
502	non-insulin-dependent diabetes	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
503	non-insulin-dependent diabetes mellitus	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
504	insulin	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
505	diabetes insipidus	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
506	ketoacidosis prone diabetes	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
507	lente insulin	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
508	sugar diabetes	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
509	chemical diabetes	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
510	nephrogenic diabetes insipidus	ketoacidosis-resistant diabetes mellitus	34	583	0.058319
511	hypoglycemic agent	ketoacidosis-resistant diabetes mellitus	33	583	0.056604
512	bronzed diabetes	ketoacidosis-resistant diabetes mellitus	33	583	0.056604
513	insulin reaction	ketoacidosis-resistant diabetes mellitus	33	583	0.056604
514	insulin shock	ketoacidosis-resistant diabetes mellitus	65	583	0.111492
515	insulin shock therapy	ketoacidosis-resistant diabetes mellitus	33	583	0.056604

516	obesity diet	ketoacidosis-resistant diabetes mellitus	33	583	0.056604
517	soft diet	ketoacidosis-resistant diabetes mellitus	33	583	0.056604
518	bland diet	ketoacidosis-resistant diabetes mellitus	33	583	0.056604
519	diabetes	ketoacidosis-resistant diabetes mellitus	33	583	0.056604
520	recombinant human insulin	ketoacidosis-resistant diabetes mellitus	33	583	0.056604
521	adult-onset diabetes	ketosis-resistant diabetes	34	583	0.058319
522	diabetes mellitus	ketosis-resistant diabetes	34	583	0.058319
523	gestational diabetes	ketosis-resistant diabetes	55	583	0.09434
524	juvenile diabetes	ketosis-resistant diabetes	34	583	0.058319
525	adult-onset diabetes mellitus	ketosis-resistant diabetes	34	583	0.058319
526	autoimmune diabetes	ketosis-resistant diabetes	34	583	0.058319
527	insulin-dependent diabetes mellitus	ketosis-resistant diabetes	34	583	0.058319
528	ketoacidosis-prone diabetes	ketosis-resistant diabetes	34	583	0.058319
529	ketoacidosis-resistant diabetes	ketosis-resistant diabetes	34	583	0.058319
530	ketoacidosis-resistant diabetes mellitus	ketosis-resistant diabetes	34	583	0.058319
531	ketosis-prone diabetes	ketosis-resistant diabetes	34	583	0.058319
532	ketosis-resistant diabetes mellitus	ketosis-resistant diabetes	34	583	0.058319
533	mature-onset diabetes	ketosis-resistant diabetes	34	583	0.058319
534	maturity-onset diabetes mellitus	ketosis-resistant diabetes	34	583	0.058319
535	non-insulin-dependent diabetes	ketosis-resistant diabetes	34	583	0.058319

536	non-insulin-dependent diabetes mellitus	ketosis-resistant diabetes	34	583	0.058319
537	insulin	ketosis-resistant diabetes	34	583	0.058319
538	diabetes insipidus	ketosis-resistant diabetes	34	583	0.058319
539	ketoacidosis prone diabetes	ketosis-resistant diabetes	34	583	0.058319
540	lente insulin	ketosis-resistant diabetes	34	583	0.058319
541	sugar diabetes	ketosis-resistant diabetes	34	583	0.058319
542	chemical diabetes	ketosis-resistant diabetes	34	583	0.058319
543	nephrogenic diabetes insipidus	ketosis-resistant diabetes	34	583	0.058319
544	hypoglycemic agent	ketosis-resistant diabetes	33	583	0.056604
545	bronzed diabetes	ketosis-resistant diabetes	33	583	0.056604
546	insulin reaction	ketosis-resistant diabetes	33	583	0.056604
547	insulin shock	ketosis-resistant diabetes	65	583	0.111492
548	insulin shock therapy	ketosis-resistant diabetes	33	583	0.056604
549	obesity diet	ketosis-resistant diabetes	33	583	0.056604
550	soft diet	ketosis-resistant diabetes	33	583	0.056604
551	bland diet	ketosis-resistant diabetes	33	583	0.056604
552	diabetes	ketosis-resistant diabetes	33	583	0.056604
553	recombinant human insulin	ketosis-resistant diabetes	33	583	0.056604
554	adult-onset diabetes	ketosis-prone diabetes	33	568	0.058099
555	diabetes mellitus	ketosis-prone diabetes	34	568	0.059859
556	gestational diabetes	ketosis-prone diabetes	57	568	0.100352
557	juvenile diabetes	ketosis-prone diabetes	34	568	0.059859
558	adult-onset diabetes mellitus	ketosis-prone diabetes	34	568	0.059859

559	autoimmune diabetes	ketosis-prone diabetes	34	568	0.059859
560	growth-onset diabetes	ketosis-prone diabetes	33	568	0.058099
561	insulin-dependent diabetes mellitus	ketosis-prone diabetes	34	568	0.059859
562	ketoacidosis-prone diabetes	ketosis-prone diabetes	33	568	0.058099
563	ketoacidosis-resistant diabetes	ketosis-prone diabetes	33	568	0.058099
564	ketoacidosis-resistant diabetes mellitus	ketosis-prone diabetes	33	568	0.058099
565	ketosis-resistant diabetes	ketosis-prone diabetes	34	568	0.059859
566	ketosis-resistant diabetes mellitus	ketosis-prone diabetes	34	568	0.059859
567	mature-onset diabetes	ketosis-prone diabetes	33	568	0.058099
568	maturity-onset diabetes	ketosis-prone diabetes	33	568	0.058099
569	maturity-onset diabetes mellitus	ketosis-prone diabetes	34	568	0.059859
570	non-insulin-dependent diabetes	ketosis-prone diabetes	32	568	0.056338
571	non-insulin-dependent diabetes mellitus	ketosis-prone diabetes	34	568	0.059859
572	insulin	ketosis-prone diabetes	32	568	0.056338
573	diabetes insipidus	ketosis-prone diabetes	34	568	0.059859
574	ketoacidosis prone diabetes	ketosis-prone diabetes	33	568	0.058099
575	lente insulin	ketosis-prone diabetes	32	568	0.056338
576	sugar diabetes	ketosis-prone diabetes	33	568	0.058099
577	chemical diabetes	ketosis-prone diabetes	34	568	0.059859
578	latent diabetes	ketosis-prone diabetes	32	568	0.056338

579	nephrogenic diabetes insipidus	ketosis-prone diabetes	34	568	0.059859
580	bronzed diabetes	ketosis-prone diabetes	33	568	0.058099
581	ketoacidosis	ketosis-prone diabetes	20	568	0.035211
582	polyuria	ketosis-prone diabetes	28	568	0.049296
583	insulin reaction	ketosis-prone diabetes	33	568	0.058099
584	insulin shock	ketosis-prone diabetes	65	568	0.114437
585	diabetes	ketosis-prone diabetes	33	568	0.058099
586	recombinant human insulin	ketosis-prone diabetes	33	568	0.058099
587	adult-onset diabetes	ketosis-resistant diabetes mellitus	33	551	0.059891
588	diabetes mellitus	ketosis-resistant diabetes mellitus	33	551	0.059891
589	gestational diabetes	ketosis-resistant diabetes mellitus	54	551	0.098004
590	juvenile diabetes	ketosis-resistant diabetes mellitus	33	551	0.059891
591	adult-onset diabetes mellitus	ketosis-resistant diabetes mellitus	33	551	0.059891
592	autoimmune diabetes	ketosis-resistant diabetes mellitus	33	551	0.059891
593	insulin-dependent diabetes mellitus	ketosis-resistant diabetes mellitus	33	551	0.059891
594	ketoacidosis-prone diabetes	ketosis-resistant diabetes mellitus	33	551	0.059891
595	ketoacidosis-resistant diabetes	ketosis-resistant diabetes mellitus	33	551	0.059891
596	ketoacidosis-resistant diabetes mellitus	ketosis-resistant diabetes mellitus	33	551	0.059891
597	ketosis-resistant diabetes	ketosis-resistant diabetes mellitus	33	551	0.059891
598	ketosis-prone diabetes	ketosis-resistant diabetes mellitus	33	551	0.059891
599	mature-onset diabetes	ketosis-resistant diabetes mellitus	33	551	0.059891
600	maturity-onset diabetes mellitus	ketosis-resistant diabetes mellitus	33	551	0.059891

601	non-insulin-dependent diabetes	ketosis-resistant diabetes mellitus	33	551	0.059891
602	non-insulin-dependent diabetes mellitus	ketosis-resistant diabetes mellitus	33	551	0.059891
603	insulin	ketosis-resistant diabetes mellitus	33	551	0.059891
604	diabetes insipidus	ketosis-resistant diabetes mellitus	33	551	0.059891
605	ketoacidosis prone diabetes	ketosis-resistant diabetes mellitus	33	551	0.059891
606	lente insulin	ketosis-resistant diabetes mellitus	33	551	0.059891
607	sugar diabetes	ketosis-resistant diabetes mellitus	33	551	0.059891
608	chemical diabetes	ketosis-resistant diabetes mellitus	33	551	0.059891
609	nephrogenic diabetes insipidus	ketosis-resistant diabetes mellitus	33	551	0.059891
610	hypoglycemic agent	ketosis-resistant diabetes mellitus	33	551	0.059891
611	bronzed diabetes	ketosis-resistant diabetes mellitus	32	551	0.058076
612	insulin reaction	ketosis-resistant diabetes mellitus	32	551	0.058076
613	insulin shock	ketosis-resistant diabetes mellitus	63	551	0.114338
614	insulin shock therapy	ketosis-resistant diabetes mellitus	32	551	0.058076
615	obesity diet	ketosis-resistant diabetes mellitus	32	551	0.058076
616	soft diet	ketosis-resistant diabetes mellitus	1	551	0.001815
617	bland diet	ketosis-resistant diabetes mellitus	32	551	0.058076
618	diabetes	ketosis-resistant diabetes mellitus	32	551	0.058076
619	recombinant human insulin	ketosis-resistant diabetes mellitus	32	551	0.058076
620	adult-onset diabetes	mature-onset diabetes	33	550	0.06
621	diabetes mellitus	mature-onset diabetes	33	550	0.06
622	gestational diabetes	mature-onset diabetes	54	550	0.098182
623	juvenile diabetes	mature-onset diabetes	33	550	0.06

624	adult-onset diabetes mellitus	mature-onset diabetes	33	550	0.06
625	autoimmune diabetes	mature-onset diabetes	33	550	0.06
626	insulin-dependent diabetes mellitus	mature-onset diabetes	33	550	0.06
627	ketoacidosis-prone diabetes	mature-onset diabetes	33	550	0.06
628	ketoacidosis-resistant diabetes	mature-onset diabetes	33	550	0.06
629	ketoacidosis-resistant diabetes mellitus	mature-onset diabetes	33	550	0.06
630	ketosis-resistant diabetes	mature-onset diabetes	33	550	0.06
631	ketosis-prone diabetes	mature-onset diabetes	33	550	0.06
632	ketosis-resistant diabetes mellitus	mature-onset diabetes	33	550	0.06
633	maturity-onset diabetes mellitus	mature-onset diabetes	33	550	0.06
634	non-insulin-dependent diabetes	mature-onset diabetes	33	550	0.06
635	non-insulin-dependent diabetes mellitus	mature-onset diabetes	33	550	0.06
636	insulin	mature-onset diabetes	33	550	0.06
637	diabetes insipidus	mature-onset diabetes	33	550	0.06
638	ketoacidosis prone diabetes	mature-onset diabetes	33	550	0.06
639	lente insulin	mature-onset diabetes	33	550	0.06
640	sugar diabetes	mature-onset diabetes	33	550	0.06
641	chemical diabetes	mature-onset diabetes	33	550	0.06
642	nephrogenic diabetes insipidus	mature-onset diabetes	33	550	0.06

643	hypoglycemic agent	mature-onset diabetes	32	550	0.058182
644	bronzed diabetes	mature-onset diabetes	32	550	0.058182
645	insulin reaction	mature-onset diabetes	32	550	0.058182
646	insulin shock	mature-onset diabetes	63	550	0.114546
647	insulin shock therapy	mature-onset diabetes	32	550	0.058182
648	obesity diet	mature-onset diabetes	32	550	0.058182
649	bland diet	mature-onset diabetes	32	550	0.058182
650	diabetes	mature-onset diabetes	32	550	0.058182
651	recombinant human insulin	mature-onset diabetes	32	550	0.058182
652	adult-onset diabetes	maturity-onset diabetes	34	584	0.058219
653	diabetes mellitus	maturity-onset diabetes	34	584	0.058219
654	gestational diabetes	maturity-onset diabetes	56	584	0.09589
655	juvenile diabetes	maturity-onset diabetes	34	584	0.058219
656	adult-onset diabetes mellitus	maturity-onset diabetes	34	584	0.058219
657	autoimmune diabetes	maturity-onset diabetes	34	584	0.058219
658	insulin-dependent diabetes mellitus	maturity-onset diabetes	34	584	0.058219
659	ketoacidosis-prone diabetes	maturity-onset diabetes	34	584	0.058219
660	ketoacidosis-resistant diabetes	maturity-onset diabetes	34	584	0.058219
661	ketoacidosis-resistant diabetes mellitus	maturity-onset diabetes	34	584	0.058219
662	ketosis-resistant diabetes	maturity-onset diabetes	34	584	0.058219
663	ketosis-prone diabetes	maturity-onset diabetes	34	584	0.058219
664	ketosis-resistant diabetes mellitus	maturity-onset diabetes	34	584	0.058219
665	mature-onset diabetes	maturity-onset diabetes	34	584	0.058219

666	maturity-onset diabetes mellitus	maturity-onset diabetes	34	584	0.058219
667	non-insulin-dependent diabetes	maturity-onset diabetes	34	584	0.058219
668	non-insulin-dependent diabetes mellitus	maturity-onset diabetes	34	584	0.058219
669	insulin	maturity-onset diabetes	34	584	0.058219
670	diabetes insipidus	maturity-onset diabetes	34	584	0.058219
671	ketoacidosis prone diabetes	maturity-onset diabetes	34	584	0.058219
672	lente insulin	maturity-onset diabetes	34	584	0.058219
673	sugar diabetes	maturity-onset diabetes	34	584	0.058219
674	chemical diabetes	maturity-onset diabetes	34	584	0.058219
675	nephrogenic diabetes insipidus	maturity-onset diabetes	34	584	0.058219
676	hypoglycemic agent	maturity-onset diabetes	33	584	0.056507
677	bronzed diabetes	maturity-onset diabetes	33	584	0.056507
678	insulin reaction	maturity-onset diabetes	33	584	0.056507
679	insulin shock	maturity-onset diabetes	65	584	0.111301
680	insulin shock therapy	maturity-onset diabetes	33	584	0.056507
681	obesity diet	maturity-onset diabetes	33	584	0.056507
682	bland diet	maturity-onset diabetes	33	584	0.056507
683	diabetes	maturity-onset diabetes	33	584	0.056507
684	recombinant human insulin	maturity-onset diabetes	33	584	0.056507
685	adult-onset diabetes	maturity-onset diabetes mellitus	33	550	0.06
686	diabetes mellitus	maturity-onset diabetes mellitus	33	550	0.06
687	gestational diabetes	maturity-onset diabetes mellitus	54	550	0.098182
688	juvenile diabetes	maturity-onset diabetes mellitus	33	550	0.06
689	adult-onset diabetes mellitus	maturity-onset diabetes mellitus	33	550	0.06
690	autoimmune diabetes	maturity-onset diabetes mellitus	33	550	0.06
691	insulin-dependent	maturity-onset diabetes mellitus	33	550	0.06

	diabetes mellitus				
692	ketoacidosis-prone diabetes	maturity-onset diabetes mellitus	33	550	0.06
693	ketoacidosis-resistant diabetes	maturity-onset diabetes mellitus	33	550	0.06
694	ketoacidosis-resistant diabetes mellitus	maturity-onset diabetes mellitus	33	550	0.06
695	ketosis-resistant diabetes	maturity-onset diabetes mellitus	33	550	0.06
696	ketosis-prone diabetes	maturity-onset diabetes mellitus	33	550	0.06
697	ketosis-resistant diabetes mellitus	maturity-onset diabetes mellitus	33	550	0.06
698	mature-onset diabetes	maturity-onset diabetes mellitus	33	550	0.06
699	non-insulin-dependent diabetes	maturity-onset diabetes mellitus	33	550	0.06
700	non-insulin-dependent diabetes mellitus	maturity-onset diabetes mellitus	33	550	0.06
701	insulin	maturity-onset diabetes mellitus	33	550	0.06
702	diabetes insipidus	maturity-onset diabetes mellitus	33	550	0.06
703	ketoacidosis prone diabetes	maturity-onset diabetes mellitus	33	550	0.06
704	lente insulin	maturity-onset diabetes mellitus	33	550	0.06
705	sugar diabetes	maturity-onset diabetes mellitus	33	550	0.06
706	chemical diabetes	maturity-onset diabetes mellitus	33	550	0.06
707	nephrogenic diabetes insipidus	maturity-onset diabetes mellitus	33	550	0.06
708	hypoglycemic agent	maturity-onset diabetes mellitus	32	550	0.058182
709	bronzed diabetes	maturity-onset diabetes mellitus	32	550	0.058182
710	insulin reaction	maturity-onset diabetes mellitus	32	550	0.058182

711	insulin shock	maturity-onset diabetes mellitus	63	550	0.114546
712	insulin shock therapy	maturity-onset diabetes mellitus	32	550	0.058182
713	obesity diet	maturity-onset diabetes mellitus	32	550	0.058182
714	bland diet	maturity-onset diabetes mellitus	32	550	0.058182
715	diabetes	maturity-onset diabetes mellitus	32	550	0.058182
716	recombinant human insulin	maturity-onset diabetes mellitus	32	550	0.058182
717	adult-onset diabetes	niddm	34	584	0.058219
718	diabetes mellitus	niddm	34	584	0.058219
719	gestational diabetes	niddm	56	584	0.09589
720	juvenile diabetes	niddm	34	584	0.058219
721	adult-onset diabetes mellitus	niddm	34	584	0.058219
722	autoimmune diabetes	niddm	34	584	0.058219
723	insulin-dependent diabetes mellitus	niddm	34	584	0.058219
724	ketoacidosis-prone diabetes	niddm	34	584	0.058219
725	ketoacidosis-resistant diabetes	niddm	34	584	0.058219
726	ketoacidosis-resistant diabetes mellitus	niddm	34	584	0.058219
727	ketosis-resistant diabetes	niddm	34	584	0.058219
728	ketosis-prone diabetes	niddm	34	584	0.058219
729	ketosis-resistant diabetes mellitus	niddm	34	584	0.058219
730	mature-onset diabetes	niddm	34	584	0.058219

731	maturity-onset diabetes mellitus	niddm	34	584	0.058219
732	non-insulin-dependent diabetes	niddm	34	584	0.058219
733	non-insulin-dependent diabetes mellitus	niddm	34	584	0.058219
734	insulin	niddm	34	584	0.058219
735	diabetes insipidus	niddm	34	584	0.058219
736	ketoacidosis prone diabetes	niddm	34	584	0.058219
737	lente insulin	niddm	34	584	0.058219
738	sugar diabetes	niddm	34	584	0.058219
739	chemical diabetes	niddm	34	584	0.058219
740	nephrogenic diabetes insipidus	niddm	34	584	0.058219
741	hypoglycemic agent	niddm	33	584	0.056507
742	bronzed diabetes	niddm	33	584	0.056507
743	insulin reaction	niddm	33	584	0.056507
744	insulin shock	niddm	65	584	0.111301
745	insulin shock therapy	niddm	33	584	0.056507
746	obesity diet	niddm	33	584	0.056507
747	bland diet	niddm	33	584	0.056507
748	diabetes	niddm	33	584	0.056507
749	recombinant human insulin	niddm	33	584	0.056507
750	adult-onset diabetes	non-insulin-dependent diabetes	33	550	0.06
751	diabetes mellitus	non-insulin-dependent diabetes	33	550	0.06
752	gestational diabetes	non-insulin-dependent diabetes	54	550	0.098182
753	juvenile diabetes	non-insulin-dependent diabetes	33	550	0.06
754	adult-onset diabetes mellitus	non-insulin-dependent diabetes	33	550	0.06
755	autoimmune diabetes	non-insulin-dependent diabetes	33	550	0.06
756	insulin-dependent	non-insulin-dependent diabetes	33	550	0.06

	diabetes mellitus				
757	ketoacidosis-prone diabetes	non-insulin-dependent diabetes	33	550	0.06
758	ketoacidosis-resistant diabetes	non-insulin-dependent diabetes	33	550	0.06
759	ketoacidosis-resistant diabetes mellitus	non-insulin-dependent diabetes	33	550	0.06
760	ketosis-resistant diabetes	non-insulin-dependent diabetes	33	550	0.06
761	ketosis-prone diabetes	non-insulin-dependent diabetes	33	550	0.06
762	ketosis-resistant diabetes mellitus	non-insulin-dependent diabetes	33	550	0.06
763	mature-onset diabetes	non-insulin-dependent diabetes	33	550	0.06
764	maturity-onset diabetes mellitus	non-insulin-dependent diabetes	33	550	0.06
765	non-insulin-dependent diabetes mellitus	non-insulin-dependent diabetes	33	550	0.06
766	insulin	non-insulin-dependent diabetes	33	550	0.06
767	diabetes insipidus	non-insulin-dependent diabetes	33	550	0.06
768	ketoacidosis prone diabetes	non-insulin-dependent diabetes	33	550	0.06
769	lente insulin	non-insulin-dependent diabetes	33	550	0.06
770	sugar diabetes	non-insulin-dependent diabetes	33	550	0.06
771	chemical diabetes	non-insulin-dependent diabetes	33	550	0.06
772	nephrogenic diabetes insipidus	non-insulin-dependent diabetes	33	550	0.06
773	hypoglycemic agent	non-insulin-dependent diabetes	32	550	0.058182
774	bronzed diabetes	non-insulin-dependent diabetes	32	550	0.058182
775	insulin reaction	non-insulin-dependent diabetes	32	550	0.058182

776	insulin shock	non-insulin-dependent diabetes	63	550	0.114546
777	insulin shock therapy	non-insulin-dependent diabetes	32	550	0.058182
778	obesity diet	non-insulin-dependent diabetes	32	550	0.058182
779	bland diet	non-insulin-dependent diabetes	32	550	0.058182
780	diabetes	non-insulin-dependent diabetes	32	550	0.058182
781	recombinant human insulin	non-insulin-dependent diabetes	32	550	0.058182
782	adult-onset diabetes	tolbutamide	34	549	0.061931
783	diabetes mellitus	tolbutamide	34	549	0.061931
784	gestational diabetes	tolbutamide	60	549	0.10929
785	juvenile diabetes	tolbutamide	34	549	0.061931
786	type i diabetes	tolbutamide	27	549	0.04918
787	adult-onset diabetes mellitus	tolbutamide	34	549	0.061931
788	autoimmune diabetes	tolbutamide	34	549	0.061931
789	growth-onset diabetes	tolbutamide	33	549	0.060109
790	insulin-dependent diabetes mellitus	tolbutamide	34	549	0.061931
791	ketoacidosis-prone diabetes	tolbutamide	34	549	0.061931
792	ketoacidosis-resistant diabetes	tolbutamide	34	549	0.061931
793	ketoacidosis-resistant diabetes mellitus	tolbutamide	34	549	0.061931
794	ketosis-resistant diabetes	tolbutamide	34	549	0.061931
795	ketosis-prone diabetes	tolbutamide	34	549	0.061931
796	ketosis-resistant diabetes mellitus	tolbutamide	34	549	0.061931

797	mature-onset diabetes	tolbutamide	34	549	0.061931
798	maturity-onset diabetes	tolbutamide	34	549	0.061931
799	maturity-onset diabetes mellitus	tolbutamide	34	549	0.061931
800	non-insulin-dependent diabetes mellitus	tolbutamide	34	549	0.061931
801	orinase	tolbutamide	22	549	0.040073
802	diabetes insipidus	tolbutamide	32	549	0.058288
803	antidiabetic drug	tolbutamide	28	549	0.051002
804	ketoacidosis prone diabetes	tolbutamide	34	549	0.061931
805	sugar diabetes	tolbutamide	34	549	0.061931
806	antidiabetic	tolbutamide	21	549	0.038251
807	chemical diabetes	tolbutamide	33	549	0.060109
808	latent diabetes	tolbutamide	31	549	0.056466
809	nephrogenic diabetes insipidus	tolbutamide	32	549	0.058288
810	bronzed diabetes	tolbutamide	31	549	0.056466
811	sulfonylurea	tolbutamide	26	549	0.047359
812	glucose tolerance test	tolbutamide	27	549	0.04918
813	insulin shock treatment	tolbutamide	27	549	0.04918
814	diabetes	tolbutamide	30	549	0.054645
815	glucose	tolbutamide	26	549	0.047359
816	adult-onset diabetes	dm	29	549	0.052823
817	diabetes mellitus	dm	34	549	0.061931
818	gestational diabetes	dm	58	549	0.105647
819	juvenile diabetes	dm	34	549	0.061931
820	adult-onset diabetes mellitus	dm	27	549	0.04918
821	autoimmune diabetes	dm	34	549	0.061931
822	growth-onset diabetes	dm	27	549	0.04918

823	insulin-dependent diabetes mellitus	dm	33	549	0.060109
824	ketoacidosis-prone diabetes	dm	34	549	0.061931
825	ketoacidosis-resistant diabetes	dm	34	549	0.061931
826	ketoacidosis-resistant diabetes mellitus	dm	34	549	0.061931
827	ketosis-resistant diabetes	dm	34	549	0.061931
828	ketosis-prone diabetes	dm	34	549	0.061931
829	ketosis-resistant diabetes mellitus	dm	34	549	0.061931
830	mature-onset diabetes	dm	28	549	0.051002
831	maturity-onset diabetes	dm	33	549	0.060109
832	maturity-onset diabetes mellitus	dm	25	549	0.045537
833	non-insulin-dependent diabetes	dm	33	549	0.060109
834	non-insulin-dependent diabetes mellitus	dm	31	549	0.056466
835	insulin	dm	31	549	0.056466
836	diabetes insipidus	dm	34	549	0.061931
837	ketoacidosis prone diabetes	dm	34	549	0.061931
838	lente insulin	dm	29	549	0.052823
839	sugar diabetes	dm	34	549	0.061931
840	chemical diabetes	dm	34	549	0.061931
841	latent diabetes	dm	34	549	0.061931
842	nephrogenic diabetes insipidus	dm	34	549	0.061931

843	bronzed diabetes	dm	33	549	0.060109
844	polyuria	dm	24	549	0.043716
845	insulin reaction	dm	28	549	0.051002
846	insulin shock	dm	55	549	0.100182
847	diabetes	dm	33	549	0.060109
848	recombinant human insulin	dm	30	549	0.054645
849	adult-onset diabetes	diabetic	33	535	0.061682
850	diabetes mellitus	diabetic	33	535	0.061682
851	gestational diabetes	diabetic	62	535	0.115888
852	juvenile diabetes	diabetic	34	535	0.063551
853	type i diabetes	diabetic	32	535	0.059813
854	type ii diabetes	diabetic	32	535	0.059813
855	adult-onset diabetes mellitus	diabetic	33	535	0.061682
856	autoimmune diabetes	diabetic	34	535	0.063551
857	growth-onset diabetes	diabetic	33	535	0.061682
858	insulin-dependent diabetes mellitus	diabetic	34	535	0.063551
859	ketoacidosis-prone diabetes	diabetic	34	535	0.063551
860	ketoacidosis-resistant diabetes	diabetic	34	535	0.063551
861	ketoacidosis-resistant diabetes mellitus	diabetic	34	535	0.063551
862	ketosis-resistant diabetes	diabetic	34	535	0.063551
863	ketosis-prone diabetes	diabetic	34	535	0.063551
864	ketosis-resistant diabetes mellitus	diabetic	34	535	0.063551
865	mature-onset diabetes	diabetic	33	535	0.061682

866	maturity-onset diabetes	diabetic	34	535	0.063551
867	maturity-onset diabetes mellitus	diabetic	33	535	0.061682
868	non-insulin-dependent diabetes	diabetic	28	535	0.052336
869	insulin	diabetic	22	535	0.041121
870	diabetes insipidus	diabetic	31	535	0.057944
871	ketoacidosis prone diabetes	diabetic	32	535	0.059813
872	lente insulin	diabetic	19	535	0.035514
873	sugar diabetes	diabetic	32	535	0.059813
874	chemical diabetes	diabetic	31	535	0.057944
875	latent diabetes	diabetic	31	535	0.057944
876	pancreas	diabetic	18	535	0.033645
877	nephrogenic diabetes insipidus	diabetic	31	535	0.057944
878	diabetical	diabetic	19	535	0.035514
879	bronzed diabetes	diabetic	29	535	0.054206
880	insulin shock	diabetic	54	535	0.100935
881	diabetes	diabetic	28	535	0.052336
882	adult-onset diabetes	non-insulin-dependent diabetes mellitus	28	488	0.057377
883	diabetes mellitus	non-insulin-dependent diabetes mellitus	32	488	0.065574
884	gestational diabetes	non-insulin-dependent diabetes mellitus	41	488	0.084016
885	juvenile diabetes	non-insulin-dependent diabetes mellitus	32	488	0.065574
886	adult-onset diabetes mellitus	non-insulin-dependent diabetes mellitus	30	488	0.061475
887	autoimmune diabetes	non-insulin-dependent diabetes mellitus	32	488	0.065574
888	insulin-dependent diabetes mellitus	non-insulin-dependent diabetes mellitus	30	488	0.061475
889	ketoacidosis-prone diabetes	non-insulin-dependent diabetes mellitus	32	488	0.065574
890	ketoacidosis-resistant diabetes	non-insulin-dependent diabetes mellitus	32	488	0.065574

891	ketoacidosis-resistant diabetes mellitus	non-insulin-dependent diabetes mellitus	32	488	0.065574
892	ketosis-resistant diabetes	non-insulin-dependent diabetes mellitus	32	488	0.065574
893	ketosis-prone diabetes	non-insulin-dependent diabetes mellitus	32	488	0.065574
894	ketosis-resistant diabetes mellitus	non-insulin-dependent diabetes mellitus	32	488	0.065574
895	mature-onset diabetes	non-insulin-dependent diabetes mellitus	28	488	0.057377
896	maturity-onset diabetes mellitus	non-insulin-dependent diabetes mellitus	30	488	0.061475
897	non-insulin-dependent diabetes	non-insulin-dependent diabetes mellitus	28	488	0.057377
898	insulin	non-insulin-dependent diabetes mellitus	33	488	0.067623
899	diabetes insipidus	non-insulin-dependent diabetes mellitus	30	488	0.061475
900	ketoacidosis prone diabetes	non-insulin-dependent diabetes mellitus	32	488	0.065574
901	lente insulin	non-insulin-dependent diabetes mellitus	33	488	0.067623
902	sugar diabetes	non-insulin-dependent diabetes mellitus	32	488	0.065574
903	chemical diabetes	non-insulin-dependent diabetes mellitus	32	488	0.065574
904	nephrogenic diabetes insipidus	non-insulin-dependent diabetes mellitus	30	488	0.061475
905	hypoglycemic agent	non-insulin-dependent diabetes mellitus	11	488	0.022541
906	bronzed diabetes	non-insulin-dependent diabetes mellitus	32	488	0.065574
907	insulin reaction	non-insulin-dependent diabetes mellitus	25	488	0.05123
908	insulin shock	non-insulin-dependent diabetes mellitus	55	488	0.112705
909	insulin shock therapy	non-insulin-dependent diabetes mellitus	29	488	0.059426
910	obesity diet	non-insulin-dependent diabetes mellitus	20	488	0.040984
911	bland diet	non-insulin-dependent diabetes mellitus	20	488	0.040984

912	diabetes	non-insulin-dependent diabetes mellitus	29	488	0.059426
913	recombinant human insulin	non-insulin-dependent diabetes mellitus	29	488	0.059426
914	adult-onset diabetes	acetonuria	30	481	0.06237
915	diabetes mellitus	acetonuria	31	481	0.064449
916	gestational diabetes	acetonuria	56	481	0.116424
917	juvenile diabetes	acetonuria	31	481	0.064449
918	adult-onset diabetes mellitus	acetonuria	31	481	0.064449
919	autoimmune diabetes	acetonuria	31	481	0.064449
920	growth-onset diabetes	acetonuria	29	481	0.060291
921	insulin-dependent diabetes mellitus	acetonuria	31	481	0.064449
922	ketoacidosis-prone diabetes	acetonuria	31	481	0.064449
923	ketoacidosis-resistant diabetes	acetonuria	31	481	0.064449
924	ketoacidosis-resistant diabetes mellitus	acetonuria	31	481	0.064449
925	ketosis-resistant diabetes	acetonuria	31	481	0.064449
926	ketosis-prone diabetes	acetonuria	31	481	0.064449
927	ketosis-resistant diabetes mellitus	acetonuria	31	481	0.064449
928	mature-onset diabetes	acetonuria	30	481	0.06237
929	maturity-onset diabetes	acetonuria	30	481	0.06237
930	maturity-onset diabetes mellitus	acetonuria	31	481	0.064449
931	non-insulin-dependent	acetonuria	31	481	0.064449

	diabetes mellitus				
932	ketoaciduria	acetonuria	29	481	0.060291
933	diabetes insipidus	acetonuria	31	481	0.064449
934	ketoacidosis prone diabetes	acetonuria	31	481	0.064449
935	sugar diabetes	acetonuria	31	481	0.064449
936	ketonuria	acetonuria	29	481	0.060291
937	chemical diabetes	acetonuria	31	481	0.064449
938	latent diabetes	acetonuria	28	481	0.058212
939	nephrogenic diabetes insipidus	acetonuria	31	481	0.064449
940	ketone body	acetonuria	60	481	0.12474
941	bronzed diabetes	acetonuria	30	481	0.06237
942	diabetes	acetonuria	29	481	0.060291
943	acetone	acetonuria	23	481	0.047817
944	diabetes mellitus	banting	30	473	0.063425
945	gestational diabetes	banting	50	473	0.105708
946	juvenile diabetes	banting	30	473	0.063425
947	autoimmune diabetes	banting	30	473	0.063425
948	insulin-dependent diabetes mellitus	banting	33	473	0.069767
949	ketoacidosis-prone diabetes	banting	30	473	0.063425
950	ketoacidosis-resistant diabetes	banting	30	473	0.063425
951	ketoacidosis-resistant diabetes mellitus	banting	30	473	0.063425
952	ketosis-resistant diabetes	banting	30	473	0.063425
953	ketosis-prone diabetes	banting	30	473	0.063425
954	ketosis-resistant diabetes mellitus	banting	30	473	0.063425

955	maturity-onset diabetes	banting	22	473	0.046512
956	non-insulin-dependent diabetes	banting	32	473	0.067653
957	non-insulin-dependent diabetes mellitus	banting	30	473	0.063425
958	insulin	banting	30	473	0.063425
959	diabetes insipidus	banting	30	473	0.063425
960	ketoacidosis prone diabetes	banting	30	473	0.063425
961	lente insulin	banting	30	473	0.063425
962	sugar diabetes	banting	30	473	0.063425
963	chemical diabetes	banting	30	473	0.063425
964	latent diabetes	banting	30	473	0.063425
965	nephrogenic diabetes insipidus	banting	30	473	0.063425
966	bronzed diabetes	banting	29	473	0.061311
967	insulin reaction	banting	31	473	0.065539
968	insulin shock	banting	60	473	0.12685
969	insulin shock therapy	banting	27	473	0.057082
970	insulin shock treatment	banting	31	473	0.065539
971	macleod	banting	12	473	0.02537
972	diabetes	banting	28	473	0.059197
973	john james rickard macleod	banting	10	473	0.021142
974	recombinant human insulin	banting	30	473	0.063425
975	john macleod	banting	10	473	0.021142
976	adult-onset diabetes	acetonemia	29	431	0.067285
977	diabetes mellitus	acetonemia	29	431	0.067285
978	gestational diabetes	acetonemia	53	431	0.12297
979	juvenile diabetes	acetonemia	29	431	0.067285
980	adult-onset diabetes mellitus	acetonemia	29	431	0.067285

981	autoimmune diabetes	acetonemia	29	431	0.067285
982	growth-onset diabetes	acetonemia	29	431	0.067285
983	insulin-dependent diabetes mellitus	acetonemia	29	431	0.067285
984	ketoacidosis-prone diabetes	acetonemia	29	431	0.067285
985	ketoacidosis-resistant diabetes	acetonemia	29	431	0.067285
986	ketoacidosis-resistant diabetes mellitus	acetonemia	29	431	0.067285
987	ketosis-resistant diabetes	acetonemia	29	431	0.067285
988	ketosis-prone diabetes	acetonemia	29	431	0.067285
989	ketosis-resistant diabetes mellitus	acetonemia	29	431	0.067285
990	mature-onset diabetes	acetonemia	29	431	0.067285
991	maturity-onset diabetes	acetonemia	29	431	0.067285
992	maturity-onset diabetes mellitus	acetonemia	29	431	0.067285
993	non-insulin-dependent diabetes mellitus	acetonemia	29	431	0.067285
994	ketonemia	acetonemia	29	431	0.067285
995	diabetes insipidus	acetonemia	29	431	0.067285
996	ketoacidosis prone diabetes	acetonemia	29	431	0.067285
997	sugar diabetes	acetonemia	29	431	0.067285
998	chemical diabetes	acetonemia	29	431	0.067285
999	nephrogenic diabetes insipidus	acetonemia	29	431	0.067285
1000	ketosis	acetonemia	29	431	0.067285
1001	ketone body	acetonemia	56	431	0.12993

1002	bronzed diabetes	acetonemia	28	431	0.064965
1003	diabetes	acetonemia	28	431	0.064965
1004	adult-onset diabetes	ketoaciduria	29	425	0.068235
1005	diabetes mellitus	ketoaciduria	29	425	0.068235
1006	gestational diabetes	ketoaciduria	53	425	0.124706
1007	juvenile diabetes	ketoaciduria	29	425	0.068235
1008	adult-onset diabetes mellitus	ketoaciduria	29	425	0.068235
1009	autoimmune diabetes	ketoaciduria	29	425	0.068235
1010	growth-onset diabetes	ketoaciduria	29	425	0.068235
1011	insulin-dependent diabetes mellitus	ketoaciduria	29	425	0.068235
1012	ketoacidosis-prone diabetes	ketoaciduria	29	425	0.068235
1013	ketoacidosis-resistant diabetes	ketoaciduria	29	425	0.068235
1014	ketoacidosis-resistant diabetes mellitus	ketoaciduria	29	425	0.068235
1015	ketosis-resistant diabetes	ketoaciduria	29	425	0.068235
1016	ketosis-prone diabetes	ketoaciduria	29	425	0.068235
1017	ketosis-resistant diabetes mellitus	ketoaciduria	29	425	0.068235
1018	mature-onset diabetes	ketoaciduria	29	425	0.068235
1019	maturity-onset diabetes	ketoaciduria	29	425	0.068235
1020	maturity-onset diabetes mellitus	ketoaciduria	29	425	0.068235
1021	non-insulin-dependent diabetes mellitus	ketoaciduria	29	425	0.068235

1022	acetonuria	ketoaciduria	29	425	0.068235
1023	diabetes insipidus	ketoaciduria	28	425	0.065882
1024	ketoacidosis prone diabetes	ketoaciduria	29	425	0.068235
1025	sugar diabetes	ketoaciduria	29	425	0.068235
1026	ketonuria	ketoaciduria	23	425	0.054118
1027	chemical diabetes	ketoaciduria	29	425	0.068235
1028	nephrogenic diabetes insipidus	ketoaciduria	28	425	0.065882
1029	ketone body	ketoaciduria	54	425	0.127059
1030	bronzed diabetes	ketoaciduria	27	425	0.063529
1031	diabetes	ketoaciduria	27	425	0.063529
1032	adult-onset diabetes	ketonemia	29	417	0.069544
1033	diabetes mellitus	ketonemia	29	417	0.069544
1034	gestational diabetes	ketonemia	53	417	0.127098
1035	juvenile diabetes	ketonemia	29	417	0.069544
1036	adult-onset diabetes mellitus	ketonemia	29	417	0.069544
1037	autoimmune diabetes	ketonemia	29	417	0.069544
1038	growth-onset diabetes	ketonemia	27	417	0.064748
1039	insulin-dependent diabetes mellitus	ketonemia	29	417	0.069544
1040	ketoacidosis-prone diabetes	ketonemia	29	417	0.069544
1041	ketoacidosis-resistant diabetes	ketonemia	29	417	0.069544
1042	ketoacidosis-resistant diabetes mellitus	ketonemia	29	417	0.069544
1043	ketosis-resistant diabetes	ketonemia	28	417	0.067146
1044	ketosis-prone diabetes	ketonemia	28	417	0.067146

1045	ketosis-resistant diabetes mellitus	ketonemia	28	417	0.067146
1046	mature-onset diabetes	ketonemia	29	417	0.069544
1047	maturity-onset diabetes	ketonemia	29	417	0.069544
1048	maturity-onset diabetes mellitus	ketonemia	29	417	0.069544
1049	non-insulin-dependent diabetes mellitus	ketonemia	29	417	0.069544
1050	acetonemia	ketonemia	25	417	0.059952
1051	diabetes insipidus	ketonemia	28	417	0.067146
1052	ketoacidosis prone diabetes	ketonemia	29	417	0.069544
1053	sugar diabetes	ketonemia	29	417	0.069544
1054	chemical diabetes	ketonemia	29	417	0.069544
1055	nephrogenic diabetes insipidus	ketonemia	28	417	0.067146
1056	ketosis	ketonemia	22	417	0.052758
1057	ketone body	ketonemia	51	417	0.122302
1058	bronzed diabetes	ketonemia	26	417	0.06235
1059	diabetes	ketonemia	25	417	0.059952
1060	adult-onset diabetes	insulin	25	402	0.062189
1061	diabetes mellitus	insulin	28	402	0.069652
1062	gestational diabetes	insulin	49	402	0.121891
1063	juvenile diabetes	insulin	29	402	0.072139
1064	adult-onset diabetes mellitus	insulin	28	402	0.069652
1065	autoimmune diabetes	insulin	29	402	0.072139
1066	ketoacidosis-prone diabetes	insulin	29	402	0.072139
1067	ketoacidosis-resistant diabetes	insulin	29	402	0.072139

1068	ketoacidosis-resistant diabetes mellitus	insulin	29	402	0.072139
1069	ketosis-resistant diabetes	insulin	29	402	0.072139
1070	ketosis-prone diabetes	insulin	29	402	0.072139
1071	ketosis-resistant diabetes mellitus	insulin	29	402	0.072139
1072	mature-onset diabetes	insulin	12	402	0.029851
1073	maturity-onset diabetes	insulin	25	402	0.062189
1074	maturity-onset diabetes mellitus	insulin	26	402	0.064677
1075	diabetes insipidus	insulin	28	402	0.069652
1076	ketoacidosis prone diabetes	insulin	28	402	0.069652
1077	sugar diabetes	insulin	31	402	0.077114
1078	chemical diabetes	insulin	28	402	0.069652
1079	latent diabetes	insulin	26	402	0.064677
1080	pancreas	insulin	31	402	0.077114
1081	nephrogenic diabetes insipidus	insulin	28	402	0.069652
1082	bronzed diabetes	insulin	26	402	0.064677
1083	iron-storage disease	insulin	23	402	0.057214
1084	glucose tolerance test	insulin	23	402	0.057214
1085	islets of langerhans	insulin	22	402	0.054726
1086	islet of langerhans	insulin	5	402	0.012438
1087	islands of langerhans	insulin	5	402	0.012438
1088	isles of langerhans	insulin	6	402	0.014925
1089	diabetes	insulin	27	402	0.067164
1090	secretes	insulin	19	402	0.047264
1091	glucose	insulin	23	402	0.057214

1092	adult-onset diabetes	orinase	28	414	0.067633
1093	diabetes mellitus	orinase	29	414	0.070048
1094	gestational diabetes	orinase	51	414	0.123188
1095	juvenile diabetes	orinase	29	414	0.070048
1096	adult-onset diabetes mellitus	orinase	29	414	0.070048
1097	autoimmune diabetes	orinase	29	414	0.070048
1098	growth-onset diabetes	orinase	28	414	0.067633
1099	insulin-dependent diabetes mellitus	orinase	29	414	0.070048
1100	ketoacidosis-prone diabetes	orinase	29	414	0.070048
1101	ketoacidosis-resistant diabetes	orinase	29	414	0.070048
1102	ketoacidosis-resistant diabetes mellitus	orinase	29	414	0.070048
1103	ketosis-resistant diabetes	orinase	29	414	0.070048
1104	ketosis-prone diabetes	orinase	29	414	0.070048
1105	ketosis-resistant diabetes mellitus	orinase	29	414	0.070048
1106	mature-onset diabetes	orinase	28	414	0.067633
1107	maturity-onset diabetes	orinase	28	414	0.067633
1108	maturity-onset diabetes mellitus	orinase	29	414	0.070048
1109	non-insulin-dependent diabetes mellitus	orinase	29	414	0.070048
1110	diabetes insipidus	orinase	28	414	0.067633

1111	antidiabetic drug	orinase	28	414	0.067633
1112	ketoacidosis prone diabetes	orinase	29	414	0.070048
1113	sugar diabetes	orinase	20	414	0.048309
1114	antidiabetic	orinase	28	414	0.067633
1115	chemical diabetes	orinase	28	414	0.067633
1116	nephrogenic diabetes insipidus	orinase	28	414	0.067633
1117	bronzed diabetes	orinase	27	414	0.065217
1118	sulfonylurea	orinase	20	414	0.048309
1119	insulin shock treatment	orinase	26	414	0.062802
1120	diabetes	orinase	26	414	0.062802
1121	adult-onset diabetes	diabetes insipidus	27	366	0.07377
1122	diabetes mellitus	diabetes insipidus	26	366	0.071038
1123	gestational diabetes	diabetes insipidus	50	366	0.136612
1124	juvenile diabetes	diabetes insipidus	27	366	0.07377
1125	type i diabetes	diabetes insipidus	27	366	0.07377
1126	type ii diabetes	diabetes insipidus	27	366	0.07377
1127	adult-onset diabetes mellitus	diabetes insipidus	27	366	0.07377
1128	autoimmune diabetes	diabetes insipidus	27	366	0.07377
1129	growth-onset diabetes	diabetes insipidus	27	366	0.07377
1130	insulin-dependent diabetes mellitus	diabetes insipidus	26	366	0.071038
1131	ketoacidosis-prone diabetes	diabetes insipidus	27	366	0.07377
1132	ketoacidosis-resistant diabetes	diabetes insipidus	27	366	0.07377
1133	ketoacidosis-resistant diabetes mellitus	diabetes insipidus	27	366	0.07377
1134	ketosis-resistant diabetes	diabetes insipidus	27	366	0.07377

1135	ketosis-prone diabetes	diabetes insipidus	27	366	0.07377
1136	ketosis-resistant diabetes mellitus	diabetes insipidus	27	366	0.07377
1137	mature-onset diabetes	diabetes insipidus	27	366	0.07377
1138	maturity-onset diabetes	diabetes insipidus	27	366	0.07377
1139	maturity-onset diabetes mellitus	diabetes insipidus	27	366	0.07377
1140	non-insulin-dependent diabetes	diabetes insipidus	18	366	0.04918
1141	ketoacidosis prone diabetes	diabetes insipidus	26	366	0.071038
1142	sugar diabetes	diabetes insipidus	26	366	0.071038
1143	chemical diabetes	diabetes insipidus	26	366	0.071038
1144	latent diabetes	diabetes insipidus	26	366	0.071038
1145	nephrogenic diabetes insipidus	diabetes insipidus	26	366	0.071038
1146	bronzed diabetes	diabetes insipidus	25	366	0.068306
1147	diabetes	diabetes insipidus	25	366	0.068306
1148	adult-onset diabetes	antidiabetic drug	26	369	0.070461
1149	diabetes mellitus	antidiabetic drug	27	369	0.073171
1150	gestational diabetes	antidiabetic drug	48	369	0.130081
1151	juvenile diabetes	antidiabetic drug	27	369	0.073171
1152	adult-onset diabetes mellitus	antidiabetic drug	27	369	0.073171
1153	autoimmune diabetes	antidiabetic drug	27	369	0.073171
1154	growth-onset diabetes	antidiabetic drug	26	369	0.070461
1155	insulin-dependent diabetes mellitus	antidiabetic drug	27	369	0.073171
1156	ketoacidosis-prone diabetes	antidiabetic drug	27	369	0.073171

1157	ketoacidosis-resistant diabetes	antidiabetic drug	27	369	0.073171
1158	ketoacidosis-resistant diabetes mellitus	antidiabetic drug	27	369	0.073171
1159	ketosis-resistant diabetes	antidiabetic drug	27	369	0.073171
1160	ketosis-prone diabetes	antidiabetic drug	27	369	0.073171
1161	ketosis-resistant diabetes mellitus	antidiabetic drug	27	369	0.073171
1162	mature-onset diabetes	antidiabetic drug	26	369	0.070461
1163	maturity-onset diabetes	antidiabetic drug	26	369	0.070461
1164	maturity-onset diabetes mellitus	antidiabetic drug	27	369	0.073171
1165	non-insulin-dependent diabetes mellitus	antidiabetic drug	27	369	0.073171
1166	diabetes insipidus	antidiabetic drug	27	369	0.073171
1167	ketoacidosis prone diabetes	antidiabetic drug	27	369	0.073171
1168	sugar diabetes	antidiabetic drug	27	369	0.073171
1169	antidiabetic	antidiabetic drug	21	369	0.056911
1170	chemical diabetes	antidiabetic drug	27	369	0.073171
1171	latent diabetes	antidiabetic drug	27	369	0.073171
1172	nephrogenic diabetes insipidus	antidiabetic drug	27	369	0.073171
1173	bronzed diabetes	antidiabetic drug	26	369	0.070461
1174	diabetes	antidiabetic drug	26	369	0.070461
1175	adult-onset diabetes	tolinase	27	362	0.074586
1176	diabetes mellitus	tolinase	27	362	0.074586
1177	gestational diabetes	tolinase	49	362	0.135359
1178	juvenile diabetes	tolinase	27	362	0.074586

1179	adult-onset diabetes mellitus	tolinase	27	362	0.074586
1180	autoimmune diabetes	tolinase	27	362	0.074586
1181	growth-onset diabetes	tolinase	27	362	0.074586
1182	insulin-dependent diabetes mellitus	tolinase	27	362	0.074586
1183	ketoacidosis-prone diabetes	tolinase	27	362	0.074586
1184	ketoacidosis-resistant diabetes	tolinase	27	362	0.074586
1185	ketoacidosis-resistant diabetes mellitus	tolinase	27	362	0.074586
1186	ketosis-resistant diabetes	tolinase	27	362	0.074586
1187	ketosis-prone diabetes	tolinase	27	362	0.074586
1188	ketosis-resistant diabetes mellitus	tolinase	27	362	0.074586
1189	mature-onset diabetes	tolinase	27	362	0.074586
1190	maturity-onset diabetes	tolinase	27	362	0.074586
1191	maturity-onset diabetes mellitus	tolinase	27	362	0.074586
1192	non-insulin-dependent diabetes mellitus	tolinase	27	362	0.074586
1193	diabetes insipidus	tolinase	22	362	0.060773
1194	ketoacidosis prone diabetes	tolinase	27	362	0.074586
1195	sugar diabetes	tolinase	27	362	0.074586
1196	chemical diabetes	tolinase	27	362	0.074586
1197	tolazamide	tolinase	22	362	0.060773

1198	nephrogenic diabetes insipidus	tolinase	22	362	0.060773
1199	bronzed diabetes	tolinase	23	362	0.063536
1200	sulfonylurea	tolinase	23	362	0.063536
1201	diabetes	tolinase	23	362	0.063536
1202	adult-onset diabetes	ketoacidosis prone diabetes	26	351	0.074074
1203	diabetes mellitus	ketoacidosis prone diabetes	26	351	0.074074
1204	gestational diabetes	ketoacidosis prone diabetes	50	351	0.14245
1205	juvenile diabetes	ketoacidosis prone diabetes	27	351	0.076923
1206	type i diabetes	ketoacidosis prone diabetes	26	351	0.074074
1207	type ii diabetes	ketoacidosis prone diabetes	3	351	0.008547
1208	adult-onset diabetes mellitus	ketoacidosis prone diabetes	26	351	0.074074
1209	autoimmune diabetes	ketoacidosis prone diabetes	26	351	0.074074
1210	growth-onset diabetes	ketoacidosis prone diabetes	26	351	0.074074
1211	insulin-dependent diabetes mellitus	ketoacidosis prone diabetes	26	351	0.074074
1212	ketoacidosis-prone diabetes	ketoacidosis prone diabetes	26	351	0.074074
1213	ketoacidosis-resistant diabetes	ketoacidosis prone diabetes	26	351	0.074074
1214	ketoacidosis-resistant diabetes mellitus	ketoacidosis prone diabetes	26	351	0.074074
1215	ketosis-resistant diabetes	ketoacidosis prone diabetes	26	351	0.074074
1216	ketosis-prone diabetes	ketoacidosis prone diabetes	26	351	0.074074
1217	ketosis-resistant diabetes mellitus	ketoacidosis prone diabetes	26	351	0.074074
1218	mature-onset diabetes	ketoacidosis prone diabetes	26	351	0.074074

1219	maturity-onset diabetes	ketoacidosis prone diabetes	26	351	0.074074
1220	maturity-onset diabetes mellitus	ketoacidosis prone diabetes	26	351	0.074074
1221	non-insulin-dependent diabetes mellitus	ketoacidosis prone diabetes	26	351	0.074074
1222	diabetes insipidus	ketoacidosis prone diabetes	26	351	0.074074
1223	sugar diabetes	ketoacidosis prone diabetes	26	351	0.074074
1224	chemical diabetes	ketoacidosis prone diabetes	26	351	0.074074
1225	latent diabetes	ketoacidosis prone diabetes	26	351	0.074074
1226	nephrogenic diabetes insipidus	ketoacidosis prone diabetes	26	351	0.074074
1227	bronzed diabetes	ketoacidosis prone diabetes	25	351	0.071225
1228	diabetes	ketoacidosis prone diabetes	25	351	0.071225
1229	diabetes mellitus	lente iletin	27	378	0.071429
1230	juvenile diabetes	lente iletin	27	378	0.071429
1231	adult-onset diabetes mellitus	lente iletin	27	378	0.071429
1232	autoimmune diabetes	lente iletin	27	378	0.071429
1233	insulin-dependent diabetes mellitus	lente iletin	27	378	0.071429
1234	ketoacidosis-prone diabetes	lente iletin	27	378	0.071429
1235	ketoacidosis-resistant diabetes	lente iletin	27	378	0.071429
1236	ketoacidosis-resistant diabetes mellitus	lente iletin	27	378	0.071429
1237	ketosis-resistant diabetes	lente iletin	27	378	0.071429
1238	ketosis-prone diabetes	lente iletin	27	378	0.071429

1239	ketosis-resistant diabetes mellitus	lente iletin	27	378	0.071429
1240	maturity-onset diabetes mellitus	lente iletin	27	378	0.071429
1241	non-insulin-dependent diabetes	lente iletin	27	378	0.071429
1242	non-insulin-dependent diabetes mellitus	lente iletin	27	378	0.071429
1243	insulin	lente iletin	27	378	0.071429
1244	diabetes insipidus	lente iletin	27	378	0.071429
1245	ketoacidosis prone diabetes	lente iletin	27	378	0.071429
1246	lente insulin	lente iletin	27	378	0.071429
1247	sugar diabetes	lente iletin	27	378	0.071429
1248	chemical diabetes	lente iletin	27	378	0.071429
1249	nephrogenic diabetes insipidus	lente iletin	27	378	0.071429
1250	bronzed diabetes	lente iletin	27	378	0.071429
1251	insulin reaction	lente iletin	27	378	0.071429
1252	insulin shock	lente iletin	53	378	0.140212
1253	insulin shock treatment	lente iletin	27	378	0.071429
1254	diabetes	lente iletin	27	378	0.071429
1255	recombinant human insulin	lente iletin	27	378	0.071429
1256	diabetes mellitus	lente insulin	26	351	0.074074
1257	juvenile diabetes	lente insulin	26	351	0.074074
1258	adult-onset diabetes mellitus	lente insulin	26	351	0.074074
1259	autoimmune diabetes	lente insulin	26	351	0.074074
1260	insulin-dependent diabetes mellitus	lente insulin	26	351	0.074074
1261	ketoacidosis-prone diabetes	lente insulin	26	351	0.074074

1262	ketoacidosis-resistant diabetes	lente insulin	26	351	0.074074
1263	ketoacidosis-resistant diabetes mellitus	lente insulin	26	351	0.074074
1264	ketosis-resistant diabetes	lente insulin	26	351	0.074074
1265	ketosis-prone diabetes	lente insulin	26	351	0.074074
1266	ketosis-resistant diabetes mellitus	lente insulin	26	351	0.074074
1267	maturity-onset diabetes mellitus	lente insulin	26	351	0.074074
1268	non-insulin-dependent diabetes	lente insulin	26	351	0.074074
1269	non-insulin-dependent diabetes mellitus	lente insulin	26	351	0.074074
1270	insulin	lente insulin	26	351	0.074074
1271	diabetes insipidus	lente insulin	26	351	0.074074
1272	ketoacidosis prone diabetes	lente insulin	26	351	0.074074
1273	sugar diabetes	lente insulin	26	351	0.074074
1274	chemical diabetes	lente insulin	26	351	0.074074
1275	nephrogenic diabetes insipidus	lente insulin	26	351	0.074074
1276	bronzed diabetes	lente insulin	26	351	0.074074
1277	insulin reaction	lente insulin	26	351	0.074074
1278	insulin shock	lente insulin	51	351	0.145299
1279	insulin shock treatment	lente insulin	26	351	0.074074
1280	diabetes	lente insulin	26	351	0.074074
1281	recombinant human insulin	lente insulin	26	351	0.074074
1282	adult-onset diabetes	sugar diabetes	26	348	0.074713
1283	diabetes mellitus	sugar diabetes	26	348	0.074713

1284	gestational diabetes	sugar diabetes	48	348	0.137931
1285	juvenile diabetes	sugar diabetes	26	348	0.074713
1286	type i diabetes	sugar diabetes	26	348	0.074713
1287	adult-onset diabetes mellitus	sugar diabetes	26	348	0.074713
1288	autoimmune diabetes	sugar diabetes	26	348	0.074713
1289	growth-onset diabetes	sugar diabetes	26	348	0.074713
1290	insulin-dependent diabetes mellitus	sugar diabetes	26	348	0.074713
1291	ketoacidosis-prone diabetes	sugar diabetes	26	348	0.074713
1292	ketoacidosis-resistant diabetes	sugar diabetes	26	348	0.074713
1293	ketoacidosis-resistant diabetes mellitus	sugar diabetes	26	348	0.074713
1294	ketosis-resistant diabetes	sugar diabetes	26	348	0.074713
1295	ketosis-prone diabetes	sugar diabetes	26	348	0.074713
1296	ketosis-resistant diabetes mellitus	sugar diabetes	26	348	0.074713
1297	mature-onset diabetes	sugar diabetes	26	348	0.074713
1298	maturity-onset diabetes	sugar diabetes	26	348	0.074713
1299	maturity-onset diabetes mellitus	sugar diabetes	26	348	0.074713
1300	non-insulin-dependent diabetes mellitus	sugar diabetes	26	348	0.074713
1301	diabetes insipidus	sugar diabetes	26	348	0.074713
1302	ketoacidosis prone diabetes	sugar diabetes	26	348	0.074713
1303	chemical diabetes	sugar diabetes	26	348	0.074713

1304	latent diabetes	sugar diabetes	26	348	0.074713
1305	nephrogenic diabetes insipidus	sugar diabetes	26	348	0.074713
1306	bronzed diabetes	sugar diabetes	25	348	0.071839
1307	diabetes	sugar diabetes	25	348	0.071839
1308	adult-onset diabetes	antidiabetic	26	348	0.074713
1309	diabetes mellitus	antidiabetic	26	348	0.074713
1310	gestational diabetes	antidiabetic	48	348	0.137931
1311	juvenile diabetes	antidiabetic	26	348	0.074713
1312	adult-onset diabetes mellitus	antidiabetic	26	348	0.074713
1313	autoimmune diabetes	antidiabetic	26	348	0.074713
1314	growth-onset diabetes	antidiabetic	26	348	0.074713
1315	insulin-dependent diabetes mellitus	antidiabetic	26	348	0.074713
1316	ketoacidosis-prone diabetes	antidiabetic	26	348	0.074713
1317	ketoacidosis-resistant diabetes	antidiabetic	26	348	0.074713
1318	ketoacidosis-resistant diabetes mellitus	antidiabetic	26	348	0.074713
1319	ketosis-resistant diabetes	antidiabetic	26	348	0.074713
1320	ketosis-prone diabetes	antidiabetic	26	348	0.074713
1321	ketosis-resistant diabetes mellitus	antidiabetic	26	348	0.074713
1322	mature-onset diabetes	antidiabetic	26	348	0.074713
1323	maturity-onset diabetes	antidiabetic	26	348	0.074713

1324	maturity-onset diabetes mellitus	antidiabetic	26	348	0.074713
1325	non-insulin-dependent diabetes mellitus	antidiabetic	26	348	0.074713
1326	diabetes insipidus	antidiabetic	26	348	0.074713
1327	ketoacidosis prone diabetes	antidiabetic	26	348	0.074713
1328	sugar diabetes	antidiabetic	26	348	0.074713
1329	chemical diabetes	antidiabetic	26	348	0.074713
1330	latent diabetes	antidiabetic	26	348	0.074713
1331	nephrogenic diabetes insipidus	antidiabetic	26	348	0.074713
1332	bronzed diabetes	antidiabetic	25	348	0.071839
1333	diabetes	antidiabetic	25	348	0.071839
1334	adult-onset diabetes	chlorpropamide	26	365	0.071233
1335	diabetes mellitus	chlorpropamide	27	365	0.073973
1336	gestational diabetes	chlorpropamide	48	365	0.131507
1337	juvenile diabetes	chlorpropamide	27	365	0.073973
1338	adult-onset diabetes mellitus	chlorpropamide	26	365	0.071233
1339	autoimmune diabetes	chlorpropamide	27	365	0.073973
1340	growth-onset diabetes	chlorpropamide	26	365	0.071233
1341	insulin-dependent diabetes mellitus	chlorpropamide	26	365	0.071233
1342	ketoacidosis-prone diabetes	chlorpropamide	27	365	0.073973
1343	ketoacidosis-resistant diabetes	chlorpropamide	27	365	0.073973
1344	ketoacidosis-resistant diabetes mellitus	chlorpropamide	27	365	0.073973

1345	ketosis-resistant diabetes	chlorpropamide	27	365	0.073973
1346	ketosis-prone diabetes	chlorpropamide	27	365	0.073973
1347	ketosis-resistant diabetes mellitus	chlorpropamide	27	365	0.073973
1348	mature-onset diabetes	chlorpropamide	26	365	0.071233
1349	maturity-onset diabetes	chlorpropamide	27	365	0.073973
1350	maturity-onset diabetes mellitus	chlorpropamide	25	365	0.068493
1351	diabetes insipidus	chlorpropamide	27	365	0.073973
1352	ketoacidosis prone diabetes	chlorpropamide	27	365	0.073973
1353	sugar diabetes	chlorpropamide	27	365	0.073973
1354	chemical diabetes	chlorpropamide	27	365	0.073973
1355	latent diabetes	chlorpropamide	27	365	0.073973
1356	nephrogenic diabetes insipidus	chlorpropamide	27	365	0.073973
1357	bronzed diabetes	chlorpropamide	26	365	0.071233
1358	insulin shock treatment	chlorpropamide	20	365	0.054795
1359	diabetes	chlorpropamide	25	365	0.068493
1360	glucose	chlorpropamide	24	365	0.065753
1361	adult-onset diabetes	biguanide	26	347	0.074928
1362	diabetes mellitus	biguanide	26	347	0.074928
1363	gestational diabetes	biguanide	47	347	0.135447
1364	juvenile diabetes	biguanide	26	347	0.074928
1365	adult-onset diabetes mellitus	biguanide	26	347	0.074928
1366	autoimmune diabetes	biguanide	26	347	0.074928
1367	growth-onset diabetes	biguanide	26	347	0.074928
1368	insulin-dependent	biguanide	26	347	0.074928

	diabetes mellitus				
1369	ketoacidosis-prone diabetes	biguanide	26	347	0.074928
1370	ketoacidosis-resistant diabetes	biguanide	26	347	0.074928
1371	ketoacidosis-resistant diabetes mellitus	biguanide	26	347	0.074928
1372	ketosis-resistant diabetes	biguanide	26	347	0.074928
1373	ketosis-prone diabetes	biguanide	26	347	0.074928
1374	ketosis-resistant diabetes mellitus	biguanide	26	347	0.074928
1375	mature-onset diabetes	biguanide	26	347	0.074928
1376	maturity-onset diabetes	biguanide	26	347	0.074928
1377	maturity-onset diabetes mellitus	biguanide	26	347	0.074928
1378	diabetes insipidus	biguanide	26	347	0.074928
1379	ketoacidosis prone diabetes	biguanide	26	347	0.074928
1380	sugar diabetes	biguanide	26	347	0.074928
1381	chemical diabetes	biguanide	26	347	0.074928
1382	latent diabetes	biguanide	26	347	0.074928
1383	nephrogenic diabetes insipidus	biguanide	26	347	0.074928
1384	bronzed diabetes	biguanide	25	347	0.072046
1385	insulin shock treatment	biguanide	25	347	0.072046
1386	diabetes	biguanide	25	347	0.072046
1387	adult-onset diabetes	ketonuria	27	354	0.076271
1388	diabetes mellitus	ketonuria	26	354	0.073446
1389	gestational diabetes	ketonuria	49	354	0.138418

1390	juvenile diabetes	ketonuria	27	354	0.076271
1391	adult-onset diabetes mellitus	ketonuria	27	354	0.076271
1392	autoimmune diabetes	ketonuria	27	354	0.076271
1393	growth-onset diabetes	ketonuria	7	354	0.019774
1394	insulin-dependent diabetes mellitus	ketonuria	26	354	0.073446
1395	ketoacidosis-prone diabetes	ketonuria	26	354	0.073446
1396	ketoacidosis-resistant diabetes	ketonuria	26	354	0.073446
1397	ketoacidosis-resistant diabetes mellitus	ketonuria	26	354	0.073446
1398	ketosis-resistant diabetes	ketonuria	26	354	0.073446
1399	ketosis-prone diabetes	ketonuria	26	354	0.073446
1400	ketosis-resistant diabetes mellitus	ketonuria	26	354	0.073446
1401	mature-onset diabetes	ketonuria	27	354	0.076271
1402	maturity-onset diabetes	ketonuria	26	354	0.073446
1403	maturity-onset diabetes mellitus	ketonuria	26	354	0.073446
1404	non-insulin-dependent diabetes mellitus	ketonuria	26	354	0.073446
1405	diabetes insipidus	ketonuria	26	354	0.073446
1406	ketoacidosis prone diabetes	ketonuria	26	354	0.073446
1407	sugar diabetes	ketonuria	26	354	0.073446
1408	chemical diabetes	ketonuria	26	354	0.073446

1409	nephrogenic diabetes insipidus	ketonuria	26	354	0.073446
1410	ketone body	ketonuria	50	354	0.141243
1411	bronzed diabetes	ketonuria	25	354	0.070621
1412	diabetes	ketonuria	25	354	0.070621
1413	adult-onset diabetes	hemochromatosis	7	348	0.020115
1414	diabetes mellitus	hemochromatosis	26	348	0.074713
1415	gestational diabetes	hemochromatosis	46	348	0.132184
1416	juvenile diabetes	hemochromatosis	27	348	0.077586
1417	adult-onset diabetes mellitus	hemochromatosis	26	348	0.074713
1418	autoimmune diabetes	hemochromatosis	26	348	0.074713
1419	growth-onset diabetes	hemochromatosis	25	348	0.071839
1420	insulin-dependent diabetes mellitus	hemochromatosis	26	348	0.074713
1421	ketoacidosis-prone diabetes	hemochromatosis	26	348	0.074713
1422	ketoacidosis-resistant diabetes	hemochromatosis	26	348	0.074713
1423	ketoacidosis-resistant diabetes mellitus	hemochromatosis	26	348	0.074713
1424	ketosis-resistant diabetes	hemochromatosis	26	348	0.074713
1425	ketosis-prone diabetes	hemochromatosis	26	348	0.074713
1426	ketosis-resistant diabetes mellitus	hemochromatosis	26	348	0.074713
1427	maturity-onset diabetes	hemochromatosis	25	348	0.071839
1428	maturity-onset diabetes mellitus	hemochromatosis	27	348	0.077586

1429	non-insulin-dependent diabetes mellitus	hemochromatosis	26	348	0.074713
1430	diabetes insipidus	hemochromatosis	26	348	0.074713
1431	ketoacidosis prone diabetes	hemochromatosis	26	348	0.074713
1432	sugar diabetes	hemochromatosis	26	348	0.074713
1433	chemical diabetes	hemochromatosis	26	348	0.074713
1434	latent diabetes	hemochromatosis	25	348	0.071839
1435	pancreas	hemochromatosis	25	348	0.071839
1436	nephrogenic diabetes insipidus	hemochromatosis	26	348	0.074713
1437	bronzed diabetes	hemochromatosis	25	348	0.071839
1438	iron-storage disease	hemochromatosis	23	348	0.066092
1439	diabetes	hemochromatosis	25	348	0.071839
1440	adult-onset diabetes	melituria	26	348	0.074713
1441	diabetes mellitus	melituria	26	348	0.074713
1442	gestational diabetes	melituria	48	348	0.137931
1443	juvenile diabetes	melituria	26	348	0.074713
1444	type i diabetes	melituria	26	348	0.074713
1445	adult-onset diabetes mellitus	melituria	26	348	0.074713
1446	autoimmune diabetes	melituria	26	348	0.074713
1447	growth-onset diabetes	melituria	26	348	0.074713
1448	insulin-dependent diabetes mellitus	melituria	26	348	0.074713
1449	ketoacidosis-prone diabetes	melituria	26	348	0.074713
1450	ketoacidosis-resistant diabetes	melituria	26	348	0.074713
1451	ketoacidosis-resistant diabetes mellitus	melituria	26	348	0.074713

1452	ketosis-resistant diabetes	melituria	26	348	0.074713
1453	ketosis-prone diabetes	melituria	26	348	0.074713
1454	ketosis-resistant diabetes mellitus	melituria	26	348	0.074713
1455	mature-onset diabetes	melituria	26	348	0.074713
1456	maturity-onset diabetes	melituria	26	348	0.074713
1457	maturity-onset diabetes mellitus	melituria	26	348	0.074713
1458	diabetes insipidus	melituria	26	348	0.074713
1459	ketoacidosis prone diabetes	melituria	26	348	0.074713
1460	sugar diabetes	melituria	26	348	0.074713
1461	chemical diabetes	melituria	26	348	0.074713
1462	latent diabetes	melituria	26	348	0.074713
1463	nephrogenic diabetes insipidus	melituria	26	348	0.074713
1464	bronzed diabetes	melituria	25	348	0.071839
1465	diabetes	melituria	25	348	0.071839
1466	adult-onset diabetes	nonketoacidosis prone	26	348	0.074713
1467	diabetes mellitus	nonketoacidosis prone	26	348	0.074713
1468	gestational diabetes	nonketoacidosis prone	48	348	0.137931
1469	juvenile diabetes	nonketoacidosis prone	26	348	0.074713
1470	adult-onset diabetes mellitus	nonketoacidosis prone	26	348	0.074713
1471	autoimmune diabetes	nonketoacidosis prone	26	348	0.074713
1472	growth-onset diabetes	nonketoacidosis prone	26	348	0.074713
1473	insulin-dependent diabetes mellitus	nonketoacidosis prone	26	348	0.074713

1474	ketoacidosis-prone diabetes	nonketoacidosis prone	26	348	0.074713
1475	ketoacidosis-resistant diabetes	nonketoacidosis prone	26	348	0.074713
1476	ketoacidosis-resistant diabetes mellitus	nonketoacidosis prone	26	348	0.074713
1477	ketosis-resistant diabetes	nonketoacidosis prone	26	348	0.074713
1478	ketosis-prone diabetes	nonketoacidosis prone	26	348	0.074713
1479	ketosis-resistant diabetes mellitus	nonketoacidosis prone	26	348	0.074713
1480	mature-onset diabetes	nonketoacidosis prone	26	348	0.074713
1481	maturity-onset diabetes	nonketoacidosis prone	26	348	0.074713
1482	maturity-onset diabetes mellitus	nonketoacidosis prone	26	348	0.074713
1483	non-insulin-dependent diabetes mellitus	nonketoacidosis prone	26	348	0.074713
1484	diabetes insipidus	nonketoacidosis prone	26	348	0.074713
1485	ketoacidosis prone diabetes	nonketoacidosis prone	26	348	0.074713
1486	sugar diabetes	nonketoacidosis prone	26	348	0.074713
1487	chemical diabetes	nonketoacidosis prone	26	348	0.074713
1488	latent diabetes	nonketoacidosis prone	26	348	0.074713
1489	nephrogenic diabetes insipidus	nonketoacidosis prone	26	348	0.074713
1490	bronzed diabetes	nonketoacidosis prone	25	348	0.071839
1491	diabetes	nonketoacidosis prone	25	348	0.071839
1492	adult-onset diabetes	chemical diabetes	25	322	0.07764
1493	diabetes mellitus	chemical diabetes	25	322	0.07764
1494	gestational diabetes	chemical diabetes	46	322	0.142857

1495	juvenile diabetes	chemical diabetes	25	322	0.07764
1496	adult-onset diabetes mellitus	chemical diabetes	25	322	0.07764
1497	autoimmune diabetes	chemical diabetes	25	322	0.07764
1498	growth-onset diabetes	chemical diabetes	25	322	0.07764
1499	insulin-dependent diabetes mellitus	chemical diabetes	25	322	0.07764
1500	ketoacidosis-prone diabetes	chemical diabetes	25	322	0.07764
1501	ketoacidosis-resistant diabetes	chemical diabetes	25	322	0.07764
1502	ketoacidosis-resistant diabetes mellitus	chemical diabetes	25	322	0.07764
1503	ketosis-resistant diabetes	chemical diabetes	25	322	0.07764
1504	ketosis-prone diabetes	chemical diabetes	25	322	0.07764
1505	ketosis-resistant diabetes mellitus	chemical diabetes	25	322	0.07764
1506	mature-onset diabetes	chemical diabetes	25	322	0.07764
1507	maturity-onset diabetes	chemical diabetes	25	322	0.07764
1508	maturity-onset diabetes mellitus	chemical diabetes	25	322	0.07764
1509	non-insulin-dependent diabetes mellitus	chemical diabetes	25	322	0.07764
1510	diabetes insipidus	chemical diabetes	25	322	0.07764
1511	ketoacidosis prone diabetes	chemical diabetes	25	322	0.07764
1512	sugar diabetes	chemical diabetes	25	322	0.07764
1513	latent diabetes	chemical diabetes	25	322	0.07764

1514	nephrogenic diabetes insipidus	chemical diabetes	25	322	0.07764
1515	bronzed diabetes	chemical diabetes	24	322	0.074534
1516	diabetes	chemical diabetes	24	322	0.074534
1517	adult-onset diabetes	latent diabetes	25	322	0.07764
1518	diabetes mellitus	latent diabetes	25	322	0.07764
1519	gestational diabetes	latent diabetes	46	322	0.142857
1520	juvenile diabetes	latent diabetes	25	322	0.07764
1521	adult-onset diabetes mellitus	latent diabetes	25	322	0.07764
1522	autoimmune diabetes	latent diabetes	25	322	0.07764
1523	growth-onset diabetes	latent diabetes	25	322	0.07764
1524	insulin-dependent diabetes mellitus	latent diabetes	25	322	0.07764
1525	ketoacidosis-prone diabetes	latent diabetes	25	322	0.07764
1526	ketoacidosis-resistant diabetes	latent diabetes	25	322	0.07764
1527	ketoacidosis-resistant diabetes mellitus	latent diabetes	25	322	0.07764
1528	ketosis-resistant diabetes	latent diabetes	25	322	0.07764
1529	ketosis-prone diabetes	latent diabetes	25	322	0.07764
1530	ketosis-resistant diabetes mellitus	latent diabetes	25	322	0.07764
1531	mature-onset diabetes	latent diabetes	25	322	0.07764
1532	maturity-onset diabetes	latent diabetes	25	322	0.07764
1533	maturity-onset diabetes mellitus	latent diabetes	25	322	0.07764

1534	non-insulin-dependent diabetes mellitus	latent diabetes	25	322	0.07764
1535	diabetes insipidus	latent diabetes	25	322	0.07764
1536	ketoacidosis prone diabetes	latent diabetes	25	322	0.07764
1537	sugar diabetes	latent diabetes	25	322	0.07764
1538	chemical diabetes	latent diabetes	25	322	0.07764
1539	nephrogenic diabetes insipidus	latent diabetes	25	322	0.07764
1540	bronzed diabetes	latent diabetes	24	322	0.074534
1541	diabetes	latent diabetes	24	322	0.074534
1542	adult-onset diabetes	tolazamide	26	347	0.074928
1543	diabetes mellitus	tolazamide	26	347	0.074928
1544	gestational diabetes	tolazamide	47	347	0.135447
1545	juvenile diabetes	tolazamide	26	347	0.074928
1546	adult-onset diabetes mellitus	tolazamide	26	347	0.074928
1547	autoimmune diabetes	tolazamide	26	347	0.074928
1548	growth-onset diabetes	tolazamide	26	347	0.074928
1549	insulin-dependent diabetes mellitus	tolazamide	26	347	0.074928
1550	ketoacidosis-prone diabetes	tolazamide	26	347	0.074928
1551	ketoacidosis-resistant diabetes	tolazamide	26	347	0.074928
1552	ketoacidosis-resistant diabetes mellitus	tolazamide	26	347	0.074928
1553	ketosis-resistant diabetes	tolazamide	26	347	0.074928
1554	ketosis-prone diabetes	tolazamide	26	347	0.074928

1555	ketosis-resistant diabetes mellitus	tolazamide	26	347	0.074928
1556	mature-onset diabetes	tolazamide	26	347	0.074928
1557	maturity-onset diabetes	tolazamide	26	347	0.074928
1558	maturity-onset diabetes mellitus	tolazamide	26	347	0.074928
1559	non-insulin-dependent diabetes mellitus	tolazamide	26	347	0.074928
1560	diabetes insipidus	tolazamide	26	347	0.074928
1561	ketoacidosis prone diabetes	tolazamide	26	347	0.074928
1562	sugar diabetes	tolazamide	26	347	0.074928
1563	chemical diabetes	tolazamide	26	347	0.074928
1564	nephrogenic diabetes insipidus	tolazamide	26	347	0.074928
1565	bronzed diabetes	tolazamide	25	347	0.072046
1566	sulfonylurea	tolazamide	25	347	0.072046
1567	diabetes	tolazamide	25	347	0.072046
1568	adult-onset diabetes	diabetic diet	25	324	0.07716
1569	diabetes mellitus	diabetic diet	24	324	0.074074
1570	gestational diabetes	diabetic diet	44	324	0.135803
1571	juvenile diabetes	diabetic diet	24	324	0.074074
1572	adult-onset diabetes mellitus	diabetic diet	23	324	0.070988
1573	autoimmune diabetes	diabetic diet	24	324	0.074074
1574	growth-onset diabetes	diabetic diet	25	324	0.07716
1575	ketoacidosis-prone diabetes	diabetic diet	24	324	0.074074
1576	ketoacidosis-resistant diabetes	diabetic diet	24	324	0.074074

1577	ketoacidosis-resistant diabetes mellitus	diabetic diet	24	324	0.074074
1578	ketosis-resistant diabetes	diabetic diet	24	324	0.074074
1579	ketosis-prone diabetes	diabetic diet	24	324	0.074074
1580	ketosis-resistant diabetes mellitus	diabetic diet	24	324	0.074074
1581	mature-onset diabetes	diabetic diet	25	324	0.07716
1582	maturity-onset diabetes	diabetic diet	27	324	0.083333
1583	maturity-onset diabetes mellitus	diabetic diet	23	324	0.070988
1584	diabetes insipidus	diabetic diet	25	324	0.07716
1585	ketoacidosis prone diabetes	diabetic diet	24	324	0.074074
1586	sugar diabetes	diabetic diet	23	324	0.070988
1587	chemical diabetes	diabetic diet	24	324	0.074074
1588	nephrogenic diabetes insipidus	diabetic diet	25	324	0.07716
1589	bronzed diabetes	diabetic diet	23	324	0.070988
1590	liquid diet	diabetic diet	7	324	0.021605
1591	fad diet	diabetic diet	7	324	0.021605
1592	obesity diet	diabetic diet	26	324	0.080247
1593	bland diet	diabetic diet	14	324	0.04321
1594	light diet	diabetic diet	6	324	0.018519
1595	allergy diet	diabetic diet	6	324	0.018519
1596	diabetes	diabetic diet	24	324	0.074074
1597	macrobiotic diet	diabetic diet	6	324	0.018519
1598	diabetes mellitus	pancreas	25	324	0.07716
1599	juvenile diabetes	pancreas	25	324	0.07716
1600	autoimmune diabetes	pancreas	25	324	0.07716
1601	insulin-dependent	pancreas	24	324	0.074074

	diabetes mellitus				
1602	ketoacidosis-prone diabetes	pancreas	25	324	0.07716
1603	ketoacidosis-resistant diabetes	pancreas	25	324	0.07716
1604	ketoacidosis-resistant diabetes mellitus	pancreas	25	324	0.07716
1605	ketosis-resistant diabetes	pancreas	25	324	0.07716
1606	ketosis-prone diabetes	pancreas	25	324	0.07716
1607	ketosis-resistant diabetes mellitus	pancreas	25	324	0.07716
1608	insulin	pancreas	27	324	0.083333
1609	diabetes insipidus	pancreas	24	324	0.074074
1610	ketoacidosis prone diabetes	pancreas	24	324	0.074074
1611	lente insulin	pancreas	28	324	0.08642
1612	sugar diabetes	pancreas	23	324	0.070988
1613	chemical diabetes	pancreas	24	324	0.074074
1614	nephrogenic diabetes insipidus	pancreas	24	324	0.074074
1615	bronzed diabetes	pancreas	24	324	0.074074
1616	somatostatin	pancreas	10	324	0.030864
1617	glucagon	pancreas	4	324	0.012346
1618	insulin reaction	pancreas	27	324	0.083333
1619	islets of langerhans	pancreas	5	324	0.015432
1620	insulin shock	pancreas	50	324	0.154321
1621	insulin shock therapy	pancreas	23	324	0.070988
1622	insulin shock treatment	pancreas	24	324	0.074074
1623	diabetes	pancreas	24	324	0.074074
1624	recombinant human insulin	pancreas	26	324	0.080247
1625	secretes	pancreas	7	324	0.021605

1626	diabetes mellitus	f	25	325	0.076923
1627	juvenile diabetes	f	25	325	0.076923
1628	autoimmune diabetes	f	25	325	0.076923
1629	insulin-dependent diabetes mellitus	f	25	325	0.076923
1630	ketoacidosis-prone diabetes	f	25	325	0.076923
1631	ketoacidosis-resistant diabetes	f	25	325	0.076923
1632	ketoacidosis-resistant diabetes mellitus	f	25	325	0.076923
1633	ketosis-resistant diabetes	f	25	325	0.076923
1634	ketosis-prone diabetes	f	25	325	0.076923
1635	ketosis-resistant diabetes mellitus	f	25	325	0.076923
1636	non-insulin-dependent diabetes	f	25	325	0.076923
1637	non-insulin-dependent diabetes mellitus	f	25	325	0.076923
1638	insulin	f	25	325	0.076923
1639	diabetes insipidus	f	25	325	0.076923
1640	ketoacidosis prone diabetes	f	25	325	0.076923
1641	lente insulin	f	25	325	0.076923
1642	sugar diabetes	f	25	325	0.076923
1643	chemical diabetes	f	25	325	0.076923
1644	nephrogenic diabetes insipidus	f	25	325	0.076923
1645	bronzed diabetes	f	25	325	0.076923
1646	insulin reaction	f	25	325	0.076923

1647	insulin shock	f	49	325	0.150769
1648	insulin shock treatment	f	25	325	0.076923
1649	diabetes	f	25	325	0.076923
1650	recombinant human insulin	f	25	325	0.076923
1651	diabetes mellitus	g	25	325	0.076923
1652	juvenile diabetes	g	25	325	0.076923
1653	autoimmune diabetes	g	25	325	0.076923
1654	insulin-dependent diabetes mellitus	g	25	325	0.076923
1655	ketoacidosis-prone diabetes	g	25	325	0.076923
1656	ketoacidosis-resistant diabetes	g	25	325	0.076923
1657	ketoacidosis-resistant diabetes mellitus	g	25	325	0.076923
1658	ketosis-resistant diabetes	g	25	325	0.076923
1659	ketosis-prone diabetes	g	25	325	0.076923
1660	ketosis-resistant diabetes mellitus	g	25	325	0.076923
1661	non-insulin-dependent diabetes	g	25	325	0.076923
1662	non-insulin-dependent diabetes mellitus	g	25	325	0.076923
1663	insulin	g	25	325	0.076923
1664	diabetes insipidus	g	25	325	0.076923
1665	ketoacidosis prone diabetes	g	25	325	0.076923
1666	lente insulin	g	25	325	0.076923
1667	sugar diabetes	g	25	325	0.076923
1668	chemical diabetes	g	25	325	0.076923

1669	nephrogenic diabetes insipidus	g	25	325	0.076923
1670	bronzed diabetes	g	25	325	0.076923
1671	insulin reaction	g	25	325	0.076923
1672	insulin shock	g	49	325	0.150769
1673	insulin shock treatment	g	25	325	0.076923
1674	diabetes	g	25	325	0.076923
1675	recombinant human insulin	g	25	325	0.076923
1676	diabetes mellitus	banting	25	325	0.076923
1677	juvenile diabetes	banting	25	325	0.076923
1678	autoimmune diabetes	banting	25	325	0.076923
1679	insulin-dependent diabetes mellitus	banting	25	325	0.076923
1680	ketoacidosis-prone diabetes	banting	25	325	0.076923
1681	ketoacidosis-resistant diabetes	banting	25	325	0.076923
1682	ketoacidosis-resistant diabetes mellitus	banting	25	325	0.076923
1683	ketosis-resistant diabetes	banting	25	325	0.076923
1684	ketosis-prone diabetes	banting	25	325	0.076923
1685	ketosis-resistant diabetes mellitus	banting	25	325	0.076923
1686	non-insulin-dependent diabetes	banting	25	325	0.076923
1687	non-insulin-dependent diabetes mellitus	banting	25	325	0.076923
1688	insulin	banting	25	325	0.076923
1689	diabetes insipidus	banting	25	325	0.076923

1690	ketoacidosis prone diabetes	banting	25	325	0.076923
1691	lente insulin	banting	25	325	0.076923
1692	sugar diabetes	banting	25	325	0.076923
1693	chemical diabetes	banting	25	325	0.076923
1694	nephrogenic diabetes insipidus	banting	25	325	0.076923
1695	bronzed diabetes	banting	25	325	0.076923
1696	insulin reaction	banting	25	325	0.076923
1697	insulin shock	banting	49	325	0.150769
1698	insulin shock treatment	banting	25	325	0.076923
1699	diabetes	banting	25	325	0.076923
1700	recombinant human insulin	banting	25	325	0.076923
1701	adult-onset diabetes	diabetologist	25	322	0.07764
1702	diabetes mellitus	diabetologist	25	322	0.07764
1703	gestational diabetes	diabetologist	46	322	0.142857
1704	juvenile diabetes	diabetologist	25	322	0.07764
1705	adult-onset diabetes mellitus	diabetologist	25	322	0.07764
1706	autoimmune diabetes	diabetologist	25	322	0.07764
1707	growth-onset diabetes	diabetologist	25	322	0.07764
1708	insulin-dependent diabetes mellitus	diabetologist	25	322	0.07764
1709	ketoacidosis-prone diabetes	diabetologist	25	322	0.07764
1710	ketoacidosis-resistant diabetes	diabetologist	25	322	0.07764
1711	ketoacidosis-resistant diabetes mellitus	diabetologist	25	322	0.07764
1712	ketosis-resistant diabetes	diabetologist	25	322	0.07764

1713	ketosis-prone diabetes	diabetologist	25	322	0.07764
1714	ketosis-resistant diabetes mellitus	diabetologist	25	322	0.07764
1715	mature-onset diabetes	diabetologist	25	322	0.07764
1716	maturity-onset diabetes	diabetologist	25	322	0.07764
1717	maturity-onset diabetes mellitus	diabetologist	25	322	0.07764
1718	diabetes insipidus	diabetologist	25	322	0.07764
1719	ketoacidosis prone diabetes	diabetologist	25	322	0.07764
1720	sugar diabetes	diabetologist	25	322	0.07764
1721	chemical diabetes	diabetologist	25	322	0.07764
1722	latent diabetes	diabetologist	25	322	0.07764
1723	nephrogenic diabetes insipidus	diabetologist	25	322	0.07764
1724	bronzed diabetes	diabetologist	24	322	0.074534
1725	diabetes	diabetologist	24	322	0.074534
1726	adult-onset diabetes	diabetophobia	25	322	0.07764
1727	diabetes mellitus	diabetophobia	25	322	0.07764
1728	gestational diabetes	diabetophobia	46	322	0.142857
1729	juvenile diabetes	diabetophobia	25	322	0.07764
1730	adult-onset diabetes mellitus	diabetophobia	25	322	0.07764
1731	autoimmune diabetes	diabetophobia	25	322	0.07764
1732	growth-onset diabetes	diabetophobia	25	322	0.07764
1733	insulin-dependent diabetes mellitus	diabetophobia	25	322	0.07764
1734	ketoacidosis-prone diabetes	diabetophobia	25	322	0.07764

1735	ketoacidosis-resistant diabetes	diabetophobia	25	322	0.07764
1736	ketoacidosis-resistant diabetes mellitus	diabetophobia	25	322	0.07764
1737	ketosis-resistant diabetes	diabetophobia	25	322	0.07764
1738	ketosis-prone diabetes	diabetophobia	25	322	0.07764
1739	ketosis-resistant diabetes mellitus	diabetophobia	25	322	0.07764
1740	mature-onset diabetes	diabetophobia	25	322	0.07764
1741	maturity-onset diabetes	diabetophobia	25	322	0.07764
1742	maturity-onset diabetes mellitus	diabetophobia	25	322	0.07764
1743	diabetes insipidus	diabetophobia	25	322	0.07764
1744	ketoacidosis prone diabetes	diabetophobia	25	322	0.07764
1745	sugar diabetes	diabetophobia	25	322	0.07764
1746	chemical diabetes	diabetophobia	25	322	0.07764
1747	latent diabetes	diabetophobia	25	322	0.07764
1748	nephrogenic diabetes insipidus	diabetophobia	25	322	0.07764
1749	bronzed diabetes	diabetophobia	24	322	0.074534
1750	diabetes	diabetophobia	24	322	0.074534
1751	diabetes mellitus	sir frederick grant banting	25	320	0.078125
1752	juvenile diabetes	sir frederick grant banting	25	320	0.078125
1753	autoimmune diabetes	sir frederick grant banting	25	320	0.078125
1754	insulin-dependent diabetes mellitus	sir frederick grant banting	25	320	0.078125
1755	ketoacidosis-prone diabetes	sir frederick grant banting	25	320	0.078125

1756	ketoacidosis-resistant diabetes	sir frederick grant banting	25	320	0.078125
1757	ketoacidosis-resistant diabetes mellitus	sir frederick grant banting	25	320	0.078125
1758	ketosis-resistant diabetes	sir frederick grant banting	25	320	0.078125
1759	ketosis-prone diabetes	sir frederick grant banting	25	320	0.078125
1760	ketosis-resistant diabetes mellitus	sir frederick grant banting	25	320	0.078125
1761	non-insulin-dependent diabetes	sir frederick grant banting	25	320	0.078125
1762	non-insulin-dependent diabetes mellitus	sir frederick grant banting	25	320	0.078125
1763	insulin	sir frederick grant banting	24	320	0.075
1764	diabetes insipidus	sir frederick grant banting	25	320	0.078125
1765	ketoacidosis prone diabetes	sir frederick grant banting	25	320	0.078125
1766	lente insulin	sir frederick grant banting	25	320	0.078125
1767	sugar diabetes	sir frederick grant banting	20	320	0.0625
1768	chemical diabetes	sir frederick grant banting	25	320	0.078125
1769	nephrogenic diabetes insipidus	sir frederick grant banting	25	320	0.078125
1770	bronzed diabetes	sir frederick grant banting	25	320	0.078125
1771	insulin reaction	sir frederick grant banting	24	320	0.075
1772	insulin shock	sir frederick grant banting	47	320	0.146875
1773	insulin shock treatment	sir frederick grant banting	25	320	0.078125
1774	diabetes	sir frederick grant banting	25	320	0.078125
1775	recombinant human insulin	sir frederick grant banting	24	320	0.075

1776	adult-onset diabetes	nephrogenic diabetes insipidus	24	288	0.083333
1777	diabetes mellitus	nephrogenic diabetes insipidus	23	288	0.079861
1778	gestational diabetes	nephrogenic diabetes insipidus	44	288	0.152778
1779	juvenile diabetes	nephrogenic diabetes insipidus	24	288	0.083333
1780	adult-onset diabetes mellitus	nephrogenic diabetes insipidus	23	288	0.079861
1781	autoimmune diabetes	nephrogenic diabetes insipidus	24	288	0.083333
1782	growth-onset diabetes	nephrogenic diabetes insipidus	24	288	0.083333
1783	insulin-dependent diabetes mellitus	nephrogenic diabetes insipidus	24	288	0.083333
1784	ketoacidosis-prone diabetes	nephrogenic diabetes insipidus	24	288	0.083333
1785	ketoacidosis-resistant diabetes	nephrogenic diabetes insipidus	24	288	0.083333
1786	ketoacidosis-resistant diabetes mellitus	nephrogenic diabetes insipidus	24	288	0.083333
1787	ketosis-resistant diabetes	nephrogenic diabetes insipidus	24	288	0.083333
1788	ketosis-prone diabetes	nephrogenic diabetes insipidus	24	288	0.083333
1789	ketosis-resistant diabetes mellitus	nephrogenic diabetes insipidus	24	288	0.083333
1790	mature-onset diabetes	nephrogenic diabetes insipidus	24	288	0.083333
1791	maturity-onset diabetes	nephrogenic diabetes insipidus	24	288	0.083333
1792	maturity-onset diabetes mellitus	nephrogenic diabetes insipidus	16	288	0.055556
1793	diabetes insipidus	nephrogenic diabetes insipidus	23	288	0.079861
1794	ketoacidosis prone diabetes	nephrogenic diabetes insipidus	23	288	0.079861
1795	sugar diabetes	nephrogenic diabetes insipidus	23	288	0.079861

1796	chemical diabetes	nephrogenic diabetes insipidus	23	288	0.079861
1797	latent diabetes	nephrogenic diabetes insipidus	22	288	0.076389
1798	bronzed diabetes	nephrogenic diabetes insipidus	22	288	0.076389
1799	diabetes	nephrogenic diabetes insipidus	22	288	0.076389
1800	diabetes mellitus	glucosuria	24	284	0.084507
1801	gestational diabetes	glucosuria	42	284	0.147887
1802	juvenile diabetes	glucosuria	24	284	0.084507
1803	adult-onset diabetes mellitus	glucosuria	23	284	0.080986
1804	autoimmune diabetes	glucosuria	24	284	0.084507
1805	insulin-dependent diabetes mellitus	glucosuria	23	284	0.080986
1806	ketoacidosis-prone diabetes	glucosuria	24	284	0.084507
1807	ketoacidosis-resistant diabetes	glucosuria	24	284	0.084507
1808	ketoacidosis-resistant diabetes mellitus	glucosuria	24	284	0.084507
1809	ketosis-resistant diabetes	glucosuria	24	284	0.084507
1810	ketosis-prone diabetes	glucosuria	24	284	0.084507
1811	ketosis-resistant diabetes mellitus	glucosuria	24	284	0.084507
1812	maturity-onset diabetes mellitus	glucosuria	23	284	0.080986
1813	non-insulin-dependent diabetes mellitus	glucosuria	23	284	0.080986
1814	diabetes insipidus	glucosuria	24	284	0.084507

1815	ketoacidosis prone diabetes	glucosuria	24	284	0.084507
1816	sugar diabetes	glucosuria	23	284	0.080986
1817	chemical diabetes	glucosuria	24	284	0.084507
1818	nephrogenic diabetes insipidus	glucosuria	24	284	0.084507
1819	bronzed diabetes	glucosuria	21	284	0.073944
1820	glucose tolerance test	glucosuria	20	284	0.070423
1821	glycosuria	glucosuria	15	284	0.052817
1822	diabetes	glucosuria	21	284	0.073944
1823	glucose	glucosuria	22	284	0.077465
1824	adult-onset diabetes	ketosis	22	287	0.076655
1825	diabetes mellitus	ketosis	24	287	0.083624
1826	gestational diabetes	ketosis	43	287	0.149826
1827	juvenile diabetes	ketosis	24	287	0.083624
1828	adult-onset diabetes mellitus	ketosis	24	287	0.083624
1829	autoimmune diabetes	ketosis	24	287	0.083624
1830	growth-onset diabetes	ketosis	26	287	0.090592
1831	insulin-dependent diabetes mellitus	ketosis	24	287	0.083624
1832	ketoacidosis-prone diabetes	ketosis	24	287	0.083624
1833	ketoacidosis-resistant diabetes	ketosis	24	287	0.083624
1834	ketoacidosis-resistant diabetes mellitus	ketosis	24	287	0.083624
1835	mature-onset diabetes	ketosis	21	287	0.073171
1836	maturity-onset diabetes	ketosis	22	287	0.076655
1837	maturity-onset diabetes mellitus	ketosis	24	287	0.083624

1838	non-insulin-dependent diabetes mellitus	ketosis	23	287	0.080139
1839	acetonemia	ketosis	1	287	0.003484
1840	ketonemia	ketosis	1	287	0.003484
1841	diabetes insipidus	ketosis	22	287	0.076655
1842	ketoacidosis prone diabetes	ketosis	24	287	0.083624
1843	sugar diabetes	ketosis	24	287	0.083624
1844	chemical diabetes	ketosis	24	287	0.083624
1845	nephrogenic diabetes insipidus	ketosis	22	287	0.076655
1846	ketone body	ketosis	36	287	0.125436
1847	bronzed diabetes	ketosis	23	287	0.080139
1848	diabetes	ketosis	23	287	0.080139
1849	diabetes mellitus	diabetic coma	22	253	0.086957
1850	juvenile diabetes	diabetic coma	22	253	0.086957
1851	type i diabetes	diabetic coma	22	253	0.086957
1852	type ii diabetes	diabetic coma	22	253	0.086957
1853	adult-onset diabetes mellitus	diabetic coma	22	253	0.086957
1854	autoimmune diabetes	diabetic coma	22	253	0.086957
1855	insulin-dependent diabetes mellitus	diabetic coma	22	253	0.086957
1856	ketoacidosis-prone diabetes	diabetic coma	22	253	0.086957
1857	ketoacidosis-resistant diabetes	diabetic coma	22	253	0.086957
1858	ketoacidosis-resistant diabetes mellitus	diabetic coma	22	253	0.086957
1859	ketosis-resistant diabetes	diabetic coma	22	253	0.086957
1860	ketosis-prone diabetes	diabetic coma	22	253	0.086957

1861	ketosis-resistant diabetes mellitus	diabetic coma	22	253	0.086957
1862	maturity-onset diabetes mellitus	diabetic coma	22	253	0.086957
1863	non-insulin-dependent diabetes mellitus	diabetic coma	22	253	0.086957
1864	diabetes insipidus	diabetic coma	22	253	0.086957
1865	ketoacidosis prone diabetes	diabetic coma	22	253	0.086957
1866	sugar diabetes	diabetic coma	22	253	0.086957
1867	chemical diabetes	diabetic coma	22	253	0.086957
1868	nephrogenic diabetes insipidus	diabetic coma	22	253	0.086957
1869	kussmaul's coma	diabetic coma	22	253	0.086957
1870	bronzed diabetes	diabetic coma	22	253	0.086957
1871	diabetes	diabetic coma	22	253	0.086957
1872	diabetes mellitus	alloxan	23	271	0.084871
1873	gestational diabetes	alloxan	42	271	0.154982
1874	juvenile diabetes	alloxan	23	271	0.084871
1875	adult-onset diabetes mellitus	alloxan	23	271	0.084871
1876	autoimmune diabetes	alloxan	23	271	0.084871
1877	insulin-dependent diabetes mellitus	alloxan	23	271	0.084871
1878	ketoacidosis-prone diabetes	alloxan	23	271	0.084871
1879	ketoacidosis-resistant diabetes	alloxan	23	271	0.084871
1880	ketoacidosis-resistant diabetes mellitus	alloxan	23	271	0.084871

1881	ketosis-resistant diabetes	alloxan	23	271	0.084871
1882	ketosis-prone diabetes	alloxan	23	271	0.084871
1883	ketosis-resistant diabetes mellitus	alloxan	23	271	0.084871
1884	maturity-onset diabetes	alloxan	21	271	0.077491
1885	maturity-onset diabetes mellitus	alloxan	22	271	0.081181
1886	non-insulin-dependent diabetes mellitus	alloxan	22	271	0.081181
1887	diabetes insipidus	alloxan	23	271	0.084871
1888	ketoacidosis prone diabetes	alloxan	23	271	0.084871
1889	sugar diabetes	alloxan	23	271	0.084871
1890	chemical diabetes	alloxan	23	271	0.084871
1891	latent diabetes	alloxan	23	271	0.084871
1892	nephrogenic diabetes insipidus	alloxan	23	271	0.084871
1893	bronzed diabetes	alloxan	22	271	0.081181
1894	diabetes	alloxan	22	271	0.081181
1895	diabetes mellitus	ketone body	21	232	0.090517
1896	gestational diabetes	ketone body	34	232	0.146552
1897	juvenile diabetes	ketone body	23	232	0.099138
1898	autoimmune diabetes	ketone body	21	232	0.090517
1899	ketoacidosis-prone diabetes	ketone body	23	232	0.099138
1900	ketoacidosis-resistant diabetes	ketone body	23	232	0.099138
1901	ketoacidosis-resistant diabetes mellitus	ketone body	21	232	0.090517

1902	ketosis-resistant diabetes	ketone body	21	232	0.090517
1903	ketosis-prone diabetes	ketone body	21	232	0.090517
1904	ketosis-resistant diabetes mellitus	ketone body	21	232	0.090517
1905	diabetes insipidus	ketone body	21	232	0.090517
1906	ketoacidosis prone diabetes	ketone body	23	232	0.099138
1907	sugar diabetes	ketone body	21	232	0.090517
1908	chemical diabetes	ketone body	21	232	0.090517
1909	latent diabetes	ketone body	18	232	0.077586
1910	nephrogenic diabetes insipidus	ketone body	21	232	0.090517
1911	bronzed diabetes	ketone body	20	232	0.086207
1912	acetoacetic acid	ketone body	6	232	0.025862
1913	glucose tolerance test	ketone body	19	232	0.081897
1914	beta cell	ketone body	8	232	0.034483
1915	diabetes	ketone body	20	232	0.086207
1916	acetone	ketone body	18	232	0.077586
1917	glucose	ketone body	19	232	0.081897
1918	diabetes mellitus	diabetical	21	258	0.081395
1919	gestational diabetes	diabetical	37	258	0.143411
1920	juvenile diabetes	diabetical	24	258	0.093023
1921	autoimmune diabetes	diabetical	24	258	0.093023
1922	insulin-dependent diabetes mellitus	diabetical	4	258	0.015504
1923	ketoacidosis-prone diabetes	diabetical	23	258	0.089147
1924	ketoacidosis-resistant diabetes	diabetical	23	258	0.089147
1925	ketoacidosis-resistant diabetes mellitus	diabetical	21	258	0.081395

1926	ketosis-resistant diabetes	diabetical	23	258	0.089147
1927	ketosis-prone diabetes	diabetical	23	258	0.089147
1928	ketosis-resistant diabetes mellitus	diabetical	21	258	0.081395
1929	maturity-onset diabetes	diabetical	19	258	0.073643
1930	diabetic	diabetical	5	258	0.01938
1931	diabetes insipidus	diabetical	23	258	0.089147
1932	ketoacidosis prone diabetes	diabetical	23	258	0.089147
1933	sugar diabetes	diabetical	24	258	0.093023
1934	chemical diabetes	diabetical	23	258	0.089147
1935	latent diabetes	diabetical	19	258	0.073643
1936	diabetic diet	diabetical	19	258	0.073643
1937	nephrogenic diabetes insipidus	diabetical	23	258	0.089147
1938	diabetic coma	diabetical	17	258	0.065891
1939	diabetic acidosis	diabetical	15	258	0.05814
1940	bronzed diabetes	diabetical	21	258	0.081395
1941	insulin shock treatment	diabetical	20	258	0.077519
1942	diabetes	diabetical	21	258	0.081395
1943	adult-onset diabetes	hyperglycemia	23	258	0.089147
1944	diabetes mellitus	hyperglycemia	22	258	0.085271
1945	gestational diabetes	hyperglycemia	41	258	0.158915
1946	juvenile diabetes	hyperglycemia	23	258	0.089147
1947	adult-onset diabetes mellitus	hyperglycemia	10	258	0.03876
1948	autoimmune diabetes	hyperglycemia	23	258	0.089147
1949	growth-onset diabetes	hyperglycemia	22	258	0.085271
1950	ketoacidosis-prone diabetes	hyperglycemia	23	258	0.089147

1951	ketoacidosis-resistant diabetes	hyperglycemia	22	258	0.085271
1952	ketoacidosis-resistant diabetes mellitus	hyperglycemia	22	258	0.085271
1953	ketosis-resistant diabetes	hyperglycemia	22	258	0.085271
1954	ketosis-prone diabetes	hyperglycemia	23	258	0.089147
1955	ketosis-resistant diabetes mellitus	hyperglycemia	22	258	0.085271
1956	mature-onset diabetes	hyperglycemia	23	258	0.089147
1957	maturity-onset diabetes	hyperglycemia	23	258	0.089147
1958	diabetes insipidus	hyperglycemia	22	258	0.085271
1959	ketoacidosis prone diabetes	hyperglycemia	22	258	0.085271
1960	sugar diabetes	hyperglycemia	22	258	0.085271
1961	chemical diabetes	hyperglycemia	22	258	0.085271
1962	nephrogenic diabetes insipidus	hyperglycemia	22	258	0.085271
1963	bronzed diabetes	hyperglycemia	21	258	0.081395
1964	diabetes	hyperglycemia	21	258	0.081395
1965	glucose	hyperglycemia	20	258	0.077519
1966	adult-onset diabetes	diuresis	23	257	0.089494
1967	diabetes mellitus	diuresis	24	257	0.093385
1968	gestational diabetes	diuresis	44	257	0.171206
1969	juvenile diabetes	diuresis	12	257	0.046693
1970	adult-onset diabetes mellitus	diuresis	22	257	0.085603
1971	autoimmune diabetes	diuresis	14	257	0.054475
1972	growth-onset diabetes	diuresis	22	257	0.085603

1973	insulin-dependent diabetes mellitus	diuresis	23	257	0.089494
1974	ketoacidosis-prone diabetes	diuresis	21	257	0.081712
1975	ketoacidosis-resistant diabetes	diuresis	21	257	0.081712
1976	ketoacidosis-resistant diabetes mellitus	diuresis	24	257	0.093385
1977	ketosis-resistant diabetes	diuresis	17	257	0.066148
1978	ketosis-prone diabetes	diuresis	16	257	0.062257
1979	ketosis-resistant diabetes mellitus	diuresis	21	257	0.081712
1980	mature-onset diabetes	diuresis	18	257	0.070039
1981	maturity-onset diabetes	diuresis	18	257	0.070039
1982	maturity-onset diabetes mellitus	diuresis	21	257	0.081712
1983	non-insulin-dependent diabetes	diuresis	18	257	0.070039
1984	non-insulin-dependent diabetes mellitus	diuresis	21	257	0.081712
1985	diabetes insipidus	diuresis	23	257	0.089494
1986	ketoacidosis prone diabetes	diuresis	24	257	0.093385
1987	nephrogenic diabetes insipidus	diuresis	23	257	0.089494
1988	bronzed diabetes	diuresis	22	257	0.085603
1989	diabetes	diuresis	22	257	0.085603
1990	diabetes mellitus	diabetic acidosis	21	231	0.090909
1991	juvenile diabetes	diabetic acidosis	21	231	0.090909

1992	adult-onset diabetes mellitus	diabetic acidosis	21	231	0.090909
1993	autoimmune diabetes	diabetic acidosis	21	231	0.090909
1994	insulin-dependent diabetes mellitus	diabetic acidosis	21	231	0.090909
1995	ketoacidosis-prone diabetes	diabetic acidosis	21	231	0.090909
1996	ketoacidosis-resistant diabetes	diabetic acidosis	21	231	0.090909
1997	ketoacidosis-resistant diabetes mellitus	diabetic acidosis	21	231	0.090909
1998	ketosis-resistant diabetes	diabetic acidosis	21	231	0.090909
1999	ketosis-prone diabetes	diabetic acidosis	21	231	0.090909
2000	ketosis-resistant diabetes mellitus	diabetic acidosis	21	231	0.090909
2001	maturity-onset diabetes mellitus	diabetic acidosis	21	231	0.090909
2002	non-insulin-dependent diabetes mellitus	diabetic acidosis	21	231	0.090909
2003	diabetes insipidus	diabetic acidosis	21	231	0.090909
2004	ketoacidosis prone diabetes	diabetic acidosis	21	231	0.090909
2005	sugar diabetes	diabetic acidosis	21	231	0.090909
2006	chemical diabetes	diabetic acidosis	21	231	0.090909
2007	nephrogenic diabetes insipidus	diabetic acidosis	21	231	0.090909
2008	ketone body	diabetic acidosis	41	231	0.177489
2009	bronzed diabetes	diabetic acidosis	21	231	0.090909
2010	diabetes	diabetic acidosis	21	231	0.090909
2011	diabetes mellitus	kussmaul's coma	22	235	0.093617

2012	juvenile diabetes	kussmaul's coma	21	235	0.089362
2013	type i diabetes	kussmaul's coma	21	235	0.089362
2014	type ii diabetes	kussmaul's coma	21	235	0.089362
2015	adult-onset diabetes mellitus	kussmaul's coma	21	235	0.089362
2016	autoimmune diabetes	kussmaul's coma	21	235	0.089362
2017	insulin-dependent diabetes mellitus	kussmaul's coma	22	235	0.093617
2018	ketoacidosis-prone diabetes	kussmaul's coma	21	235	0.089362
2019	ketoacidosis-resistant diabetes	kussmaul's coma	21	235	0.089362
2020	ketoacidosis-resistant diabetes mellitus	kussmaul's coma	22	235	0.093617
2021	ketosis-resistant diabetes	kussmaul's coma	21	235	0.089362
2022	ketosis-prone diabetes	kussmaul's coma	21	235	0.089362
2023	ketosis-resistant diabetes mellitus	kussmaul's coma	22	235	0.093617
2024	maturity-onset diabetes mellitus	kussmaul's coma	21	235	0.089362
2025	non-insulin-dependent diabetes mellitus	kussmaul's coma	21	235	0.089362
2026	diabetes insipidus	kussmaul's coma	21	235	0.089362
2027	ketoacidosis prone diabetes	kussmaul's coma	21	235	0.089362
2028	sugar diabetes	kussmaul's coma	21	235	0.089362
2029	chemical diabetes	kussmaul's coma	21	235	0.089362
2030	nephrogenic diabetes insipidus	kussmaul's coma	21	235	0.089362
2031	diabetic coma	kussmaul's coma	4	235	0.017021

2032	bronzed diabetes	kussmaul's coma	21	235	0.089362
2033	diabetes	kussmaul's coma	21	235	0.089362
2034	diabetes mellitus	hypoglycaemic agent	21	231	0.090909
2035	juvenile diabetes	hypoglycaemic agent	21	231	0.090909
2036	adult-onset diabetes mellitus	hypoglycaemic agent	21	231	0.090909
2037	autoimmune diabetes	hypoglycaemic agent	21	231	0.090909
2038	insulin-dependent diabetes mellitus	hypoglycaemic agent	21	231	0.090909
2039	ketoacidosis-prone diabetes	hypoglycaemic agent	21	231	0.090909
2040	ketoacidosis-resistant diabetes	hypoglycaemic agent	21	231	0.090909
2041	ketoacidosis-resistant diabetes mellitus	hypoglycaemic agent	21	231	0.090909
2042	ketosis-resistant diabetes	hypoglycaemic agent	21	231	0.090909
2043	ketosis-prone diabetes	hypoglycaemic agent	21	231	0.090909
2044	ketosis-resistant diabetes mellitus	hypoglycaemic agent	21	231	0.090909
2045	maturity-onset diabetes mellitus	hypoglycaemic agent	21	231	0.090909
2046	non-insulin-dependent diabetes mellitus	hypoglycaemic agent	21	231	0.090909
2047	diabetes insipidus	hypoglycaemic agent	21	231	0.090909
2048	ketoacidosis prone diabetes	hypoglycaemic agent	21	231	0.090909
2049	sugar diabetes	hypoglycaemic agent	21	231	0.090909
2050	chemical diabetes	hypoglycaemic agent	21	231	0.090909

2051	nephrogenic diabetes insipidus	hypoglycaemic agent	21	231	0.090909
2052	bronzed diabetes	hypoglycaemic agent	21	231	0.090909
2053	insulin shock treatment	hypoglycaemic agent	21	231	0.090909
2054	diabetes	hypoglycaemic agent	21	231	0.090909
2055	glucose	hypoglycaemic agent	21	231	0.090909
2056	diabetes mellitus	hypoglycemic agent	21	231	0.090909
2057	juvenile diabetes	hypoglycemic agent	21	231	0.090909
2058	adult-onset diabetes mellitus	hypoglycemic agent	21	231	0.090909
2059	autoimmune diabetes	hypoglycemic agent	21	231	0.090909
2060	insulin-dependent diabetes mellitus	hypoglycemic agent	21	231	0.090909
2061	ketoacidosis-prone diabetes	hypoglycemic agent	21	231	0.090909
2062	ketoacidosis-resistant diabetes	hypoglycemic agent	21	231	0.090909
2063	ketoacidosis-resistant diabetes mellitus	hypoglycemic agent	21	231	0.090909
2064	ketosis-resistant diabetes	hypoglycemic agent	21	231	0.090909
2065	ketosis-prone diabetes	hypoglycemic agent	21	231	0.090909
2066	ketosis-resistant diabetes mellitus	hypoglycemic agent	21	231	0.090909
2067	maturity-onset diabetes mellitus	hypoglycemic agent	21	231	0.090909
2068	non-insulin-dependent diabetes mellitus	hypoglycemic agent	21	231	0.090909
2069	diabetes insipidus	hypoglycemic agent	21	231	0.090909

2070	ketoacidosis prone diabetes	hypoglycemic agent	21	231	0.090909
2071	sugar diabetes	hypoglycemic agent	21	231	0.090909
2072	chemical diabetes	hypoglycemic agent	21	231	0.090909
2073	nephrogenic diabetes insipidus	hypoglycemic agent	21	231	0.090909
2074	bronzed diabetes	hypoglycemic agent	21	231	0.090909
2075	insulin shock treatment	hypoglycemic agent	21	231	0.090909
2076	diabetes	hypoglycemic agent	21	231	0.090909
2077	glucose	hypoglycemic agent	21	231	0.090909
2078	diabetes mellitus	bronzed diabetes	20	210	0.095238
2079	juvenile diabetes	bronzed diabetes	20	210	0.095238
2080	adult-onset diabetes mellitus	bronzed diabetes	20	210	0.095238
2081	autoimmune diabetes	bronzed diabetes	20	210	0.095238
2082	insulin-dependent diabetes mellitus	bronzed diabetes	20	210	0.095238
2083	ketoacidosis-prone diabetes	bronzed diabetes	20	210	0.095238
2084	ketoacidosis-resistant diabetes	bronzed diabetes	20	210	0.095238
2085	ketoacidosis-resistant diabetes mellitus	bronzed diabetes	20	210	0.095238
2086	ketosis-resistant diabetes	bronzed diabetes	20	210	0.095238
2087	ketosis-prone diabetes	bronzed diabetes	20	210	0.095238
2088	ketosis-resistant diabetes mellitus	bronzed diabetes	20	210	0.095238
2089	maturity-onset diabetes mellitus	bronzed diabetes	20	210	0.095238
2090	non-insulin-dependent	bronzed diabetes	20	210	0.095238

	diabetes mellitus				
2091	diabetes insipidus	bronzed diabetes	20	210	0.095238
2092	ketoacidosis prone diabetes	bronzed diabetes	20	210	0.095238
2093	sugar diabetes	bronzed diabetes	20	210	0.095238
2094	chemical diabetes	bronzed diabetes	20	210	0.095238
2095	pancreas	bronzed diabetes	20	210	0.095238
2096	nephrogenic diabetes insipidus	bronzed diabetes	20	210	0.095238
2097	iron-storage disease	bronzed diabetes	20	210	0.095238
2098	diabetes	bronzed diabetes	20	210	0.095238
2099	diabetes mellitus	iron-storage disease	20	210	0.095238
2100	juvenile diabetes	iron-storage disease	20	210	0.095238
2101	adult-onset diabetes mellitus	iron-storage disease	20	210	0.095238
2102	autoimmune diabetes	iron-storage disease	20	210	0.095238
2103	insulin-dependent diabetes mellitus	iron-storage disease	20	210	0.095238
2104	ketoacidosis-prone diabetes	iron-storage disease	20	210	0.095238
2105	ketoacidosis-resistant diabetes	iron-storage disease	20	210	0.095238
2106	ketoacidosis-resistant diabetes mellitus	iron-storage disease	20	210	0.095238
2107	ketosis-resistant diabetes	iron-storage disease	20	210	0.095238
2108	ketosis-prone diabetes	iron-storage disease	20	210	0.095238
2109	ketosis-resistant diabetes mellitus	iron-storage disease	20	210	0.095238
2110	maturity-onset diabetes mellitus	iron-storage disease	20	210	0.095238

2111	non-insulin-dependent diabetes mellitus	iron-storage disease	20	210	0.095238
2112	diabetes insipidus	iron-storage disease	20	210	0.095238
2113	ketoacidosis prone diabetes	iron-storage disease	20	210	0.095238
2114	sugar diabetes	iron-storage disease	20	210	0.095238
2115	chemical diabetes	iron-storage disease	20	210	0.095238
2116	pancreas	iron-storage disease	20	210	0.095238
2117	nephrogenic diabetes insipidus	iron-storage disease	20	210	0.095238
2118	bronzed diabetes	iron-storage disease	20	210	0.095238
2119	diabetes	iron-storage disease	20	210	0.095238
2120	adult-onset diabetes	juvenile	21	222	0.094595
2121	diabetes mellitus	juvenile	21	222	0.094595
2122	adult-onset diabetes mellitus	juvenile	20	222	0.09009
2123	autoimmune diabetes	juvenile	21	222	0.094595
2124	insulin-dependent diabetes mellitus	juvenile	18	222	0.081081
2125	ketoacidosis-prone diabetes	juvenile	21	222	0.094595
2126	ketoacidosis-resistant diabetes	juvenile	21	222	0.094595
2127	ketoacidosis-resistant diabetes mellitus	juvenile	18	222	0.081081
2128	ketosis-resistant diabetes	juvenile	21	222	0.094595
2129	ketosis-prone diabetes	juvenile	21	222	0.094595
2130	ketosis-resistant diabetes mellitus	juvenile	18	222	0.081081

2131	mature-onset diabetes	juvenile	18	222	0.081081
2132	maturity-onset diabetes	juvenile	19	222	0.085586
2133	maturity-onset diabetes mellitus	juvenile	18	222	0.081081
2134	diabetes insipidus	juvenile	21	222	0.094595
2135	ketoacidosis prone diabetes	juvenile	21	222	0.094595
2136	sugar diabetes	juvenile	21	222	0.094595
2137	chemical diabetes	juvenile	21	222	0.094595
2138	nephrogenic diabetes insipidus	juvenile	21	222	0.094595
2139	bronzed diabetes	juvenile	21	222	0.094595
2140	diabetes	juvenile	21	222	0.094595
2141	recombinant human insulin	juvenile	21	222	0.094595
2142	diabetes mellitus	sulfonylurea	22	217	0.101383
2143	juvenile diabetes	sulfonylurea	18	217	0.082949
2144	adult-onset diabetes mellitus	sulfonylurea	20	217	0.092166
2145	autoimmune diabetes	sulfonylurea	18	217	0.082949
2146	insulin-dependent diabetes mellitus	sulfonylurea	22	217	0.101383
2147	ketoacidosis-prone diabetes	sulfonylurea	18	217	0.082949
2148	ketoacidosis-resistant diabetes	sulfonylurea	18	217	0.082949
2149	ketoacidosis-resistant diabetes mellitus	sulfonylurea	22	217	0.101383
2150	ketosis-resistant diabetes	sulfonylurea	18	217	0.082949
2151	ketosis-prone diabetes	sulfonylurea	18	217	0.082949

2152	ketosis-resistant diabetes mellitus	sulfonylurea	22	217	0.101383
2153	maturity-onset diabetes mellitus	sulfonylurea	20	217	0.092166
2154	non-insulin-dependent diabetes mellitus	sulfonylurea	22	217	0.101383
2155	diabetes insipidus	sulfonylurea	21	217	0.096774
2156	antidiabetic drug	sulfonylurea	13	217	0.059908
2157	ketoacidosis prone diabetes	sulfonylurea	18	217	0.082949
2158	sugar diabetes	sulfonylurea	15	217	0.069124
2159	antidiabetic	sulfonylurea	13	217	0.059908
2160	chemical diabetes	sulfonylurea	21	217	0.096774
2161	nephrogenic diabetes insipidus	sulfonylurea	21	217	0.096774
2162	bronzed diabetes	sulfonylurea	21	217	0.096774
2163	diabetes	sulfonylurea	22	217	0.101383
2164	glucose	sulfonylurea	11	217	0.050691
2165	diabetes mellitus	iron overload	20	212	0.09434
2166	juvenile diabetes	iron overload	20	212	0.09434
2167	adult-onset diabetes mellitus	iron overload	21	212	0.099057
2168	autoimmune diabetes	iron overload	20	212	0.09434
2169	insulin-dependent diabetes mellitus	iron overload	20	212	0.09434
2170	ketoacidosis-prone diabetes	iron overload	20	212	0.09434
2171	ketoacidosis-resistant diabetes	iron overload	20	212	0.09434
2172	ketoacidosis-resistant diabetes mellitus	iron overload	20	212	0.09434

2173	ketosis-resistant diabetes	iron overload	20	212	0.09434
2174	ketosis-prone diabetes	iron overload	20	212	0.09434
2175	ketosis-resistant diabetes mellitus	iron overload	20	212	0.09434
2176	maturity-onset diabetes mellitus	iron overload	20	212	0.09434
2177	non-insulin-dependent diabetes mellitus	iron overload	20	212	0.09434
2178	diabetes insipidus	iron overload	20	212	0.09434
2179	ketoacidosis prone diabetes	iron overload	20	212	0.09434
2180	sugar diabetes	iron overload	20	212	0.09434
2181	chemical diabetes	iron overload	20	212	0.09434
2182	pancreas	iron overload	20	212	0.09434
2183	nephrogenic diabetes insipidus	iron overload	20	212	0.09434
2184	bronzed diabetes	iron overload	21	212	0.099057
2185	iron-storage disease	iron overload	2	212	0.009434
2186	diabetes	iron overload	20	212	0.09434
2187	diabetes mellitus	glucophage	19	190	0.1
2188	juvenile diabetes	glucophage	19	190	0.1
2189	type i diabetes	glucophage	19	190	0.1
2190	type ii diabetes	glucophage	19	190	0.1
2191	autoimmune diabetes	glucophage	19	190	0.1
2192	ketoacidosis-prone diabetes	glucophage	19	190	0.1
2193	ketoacidosis-resistant diabetes	glucophage	19	190	0.1
2194	ketoacidosis-resistant diabetes mellitus	glucophage	19	190	0.1

2195	ketosis-resistant diabetes	glucophage	19	190	0.1
2196	ketosis-prone diabetes	glucophage	19	190	0.1
2197	ketosis-resistant diabetes mellitus	glucophage	19	190	0.1
2198	diabetes insipidus	glucophage	19	190	0.1
2199	antidiabetic drug	glucophage	19	190	0.1
2200	ketoacidosis prone diabetes	glucophage	19	190	0.1
2201	sugar diabetes	glucophage	19	190	0.1
2202	antidiabetic	glucophage	19	190	0.1
2203	chemical diabetes	glucophage	19	190	0.1
2204	nephrogenic diabetes insipidus	glucophage	19	190	0.1
2205	bronzed diabetes	glucophage	19	190	0.1
2206	diabetes	glucophage	19	190	0.1
2207	diabetes mellitus	acetoacetic acid	20	208	0.096154
2208	juvenile diabetes	acetoacetic acid	18	208	0.086538
2209	type i diabetes	acetoacetic acid	19	208	0.091346
2210	type ii diabetes	acetoacetic acid	19	208	0.091346
2211	adult-onset diabetes mellitus	acetoacetic acid	20	208	0.096154
2212	autoimmune diabetes	acetoacetic acid	20	208	0.096154
2213	insulin-dependent diabetes mellitus	acetoacetic acid	20	208	0.096154
2214	ketoacidosis-prone diabetes	acetoacetic acid	20	208	0.096154
2215	ketoacidosis-resistant diabetes	acetoacetic acid	20	208	0.096154
2216	ketoacidosis-resistant diabetes mellitus	acetoacetic acid	20	208	0.096154

2217	ketosis-resistant diabetes	acetoacetic acid	20	208	0.096154
2218	ketosis-prone diabetes	acetoacetic acid	20	208	0.096154
2219	ketosis-resistant diabetes mellitus	acetoacetic acid	20	208	0.096154
2220	maturity-onset diabetes mellitus	acetoacetic acid	20	208	0.096154
2221	non-insulin-dependent diabetes mellitus	acetoacetic acid	20	208	0.096154
2222	diabetes insipidus	acetoacetic acid	20	208	0.096154
2223	ketoacidosis prone diabetes	acetoacetic acid	20	208	0.096154
2224	chemical diabetes	acetoacetic acid	20	208	0.096154
2225	nephrogenic diabetes insipidus	acetoacetic acid	20	208	0.096154
2226	bronzed diabetes	acetoacetic acid	20	208	0.096154
2227	diabetes	acetoacetic acid	20	208	0.096154
2228	diabetes mellitus	glucose tolerance test	20	198	0.10101
2229	juvenile diabetes	glucose tolerance test	19	198	0.09596
2230	adult-onset diabetes mellitus	glucose tolerance test	20	198	0.10101
2231	autoimmune diabetes	glucose tolerance test	19	198	0.09596
2232	insulin-dependent diabetes mellitus	glucose tolerance test	20	198	0.10101
2233	ketoacidosis-prone diabetes	glucose tolerance test	19	198	0.09596
2234	ketoacidosis-resistant diabetes	glucose tolerance test	19	198	0.09596
2235	ketoacidosis-resistant diabetes mellitus	glucose tolerance test	20	198	0.10101

2236	ketosis-resistant diabetes	glucose tolerance test	19	198	0.09596
2237	ketosis-prone diabetes	glucose tolerance test	19	198	0.09596
2238	ketosis-resistant diabetes mellitus	glucose tolerance test	20	198	0.10101
2239	maturity-onset diabetes mellitus	glucose tolerance test	20	198	0.10101
2240	non-insulin-dependent diabetes mellitus	glucose tolerance test	20	198	0.10101
2241	diabetes insipidus	glucose tolerance test	19	198	0.09596
2242	ketoacidosis prone diabetes	glucose tolerance test	19	198	0.09596
2243	sugar diabetes	glucose tolerance test	20	198	0.10101
2244	chemical diabetes	glucose tolerance test	19	198	0.09596
2245	nephrogenic diabetes insipidus	glucose tolerance test	19	198	0.09596
2246	bronzed diabetes	glucose tolerance test	19	198	0.09596
2247	hypoglycemia	glucose tolerance test	8	198	0.040404
2248	diabetes	glucose tolerance test	19	198	0.09596
2249	diabetes mellitus	somatostatin	16	206	0.07767
2250	juvenile diabetes	somatostatin	16	206	0.07767
2251	autoimmune diabetes	somatostatin	16	206	0.07767
2252	insulin-dependent diabetes mellitus	somatostatin	26	206	0.126214
2253	ketoacidosis-prone diabetes	somatostatin	16	206	0.07767
2254	ketoacidosis-resistant diabetes	somatostatin	16	206	0.07767
2255	ketoacidosis-resistant diabetes mellitus	somatostatin	13	206	0.063107

2256	ketosis-resistant diabetes	somatostatin	16	206	0.07767
2257	ketosis-prone diabetes	somatostatin	16	206	0.07767
2258	ketosis-resistant diabetes mellitus	somatostatin	7	206	0.033981
2259	non-insulin-dependent diabetes	somatostatin	24	206	0.116505
2260	non-insulin-dependent diabetes mellitus	somatostatin	24	206	0.116505
2261	insulin	somatostatin	15	206	0.072816
2262	diabetes insipidus	somatostatin	18	206	0.087379
2263	ketoacidosis prone diabetes	somatostatin	17	206	0.082524
2264	lente insulin	somatostatin	17	206	0.082524
2265	chemical diabetes	somatostatin	6	206	0.029126
2266	pancreas	somatostatin	13	206	0.063107
2267	nephrogenic diabetes insipidus	somatostatin	18	206	0.087379
2268	bronzed diabetes	somatostatin	18	206	0.087379
2269	insulin reaction	somatostatin	10	206	0.048544
2270	insulin shock	somatostatin	22	206	0.106796
2271	insulin shock therapy	somatostatin	11	206	0.053398
2272	insulin shock treatment	somatostatin	8	206	0.038835
2273	diabetes	somatostatin	22	206	0.106796
2274	recombinant human insulin	somatostatin	10	206	0.048544
2275	diabetes mellitus	ketogenesis	20	206	0.097087
2276	gestational diabetes	ketogenesis	35	206	0.169903
2277	juvenile diabetes	ketogenesis	20	206	0.097087
2278	autoimmune diabetes	ketogenesis	20	206	0.097087
2279	ketoacidosis-prone diabetes	ketogenesis	20	206	0.097087

2280	ketoacidosis-resistant diabetes	ketogenesis	20	206	0.097087
2281	ketoacidosis-resistant diabetes mellitus	ketogenesis	20	206	0.097087
2282	ketosis-resistant diabetes	ketogenesis	20	206	0.097087
2283	ketosis-prone diabetes	ketogenesis	20	206	0.097087
2284	ketosis-resistant diabetes mellitus	ketogenesis	20	206	0.097087
2285	diabetes insipidus	ketogenesis	20	206	0.097087
2286	ketoacidosis prone diabetes	ketogenesis	20	206	0.097087
2287	sugar diabetes	ketogenesis	20	206	0.097087
2288	chemical diabetes	ketogenesis	20	206	0.097087
2289	latent diabetes	ketogenesis	20	206	0.097087
2290	nephrogenic diabetes insipidus	ketogenesis	20	206	0.097087
2291	ketone body	ketogenesis	38	206	0.184466
2292	bronzed diabetes	ketogenesis	19	206	0.092233
2293	diabetes	ketogenesis	19	206	0.092233
2294	diabetes mellitus	polydipsia	19	200	0.095
2295	gestational diabetes	polydipsia	32	200	0.16
2296	juvenile diabetes	polydipsia	20	200	0.1
2297	type i diabetes	polydipsia	18	200	0.09
2298	type ii diabetes	polydipsia	18	200	0.09
2299	autoimmune diabetes	polydipsia	20	200	0.1
2300	ketoacidosis-prone diabetes	polydipsia	20	200	0.1
2301	ketoacidosis-resistant diabetes	polydipsia	20	200	0.1
2302	ketoacidosis-resistant diabetes mellitus	polydipsia	19	200	0.095

2303	ketosis-resistant diabetes	polydipsia	20	200	0.1
2304	ketosis-prone diabetes	polydipsia	20	200	0.1
2305	ketosis-resistant diabetes mellitus	polydipsia	19	200	0.095
2306	diabetes insipidus	polydipsia	20	200	0.1
2307	ketoacidosis prone diabetes	polydipsia	20	200	0.1
2308	sugar diabetes	polydipsia	20	200	0.1
2309	chemical diabetes	polydipsia	20	200	0.1
2310	latent diabetes	polydipsia	17	200	0.085
2311	nephrogenic diabetes insipidus	polydipsia	20	200	0.1
2312	bronzed diabetes	polydipsia	19	200	0.095
2313	diabetes	polydipsia	19	200	0.095
2314	diabetes mellitus	brittle diabetes	19	194	0.097938
2315	gestational diabetes	brittle diabetes	33	194	0.170103
2316	juvenile diabetes	brittle diabetes	20	194	0.103093
2317	autoimmune diabetes	brittle diabetes	20	194	0.103093
2318	ketoacidosis-prone diabetes	brittle diabetes	20	194	0.103093
2319	ketoacidosis-resistant diabetes	brittle diabetes	20	194	0.103093
2320	ketoacidosis-resistant diabetes mellitus	brittle diabetes	20	194	0.103093
2321	ketosis-resistant diabetes	brittle diabetes	20	194	0.103093
2322	ketosis-prone diabetes	brittle diabetes	20	194	0.103093
2323	ketosis-resistant diabetes mellitus	brittle diabetes	20	194	0.103093

2324	maturity-onset diabetes	brittle diabetes	8	194	0.041237
2325	diabetes insipidus	brittle diabetes	19	194	0.097938
2326	ketoacidosis prone diabetes	brittle diabetes	19	194	0.097938
2327	sugar diabetes	brittle diabetes	19	194	0.097938
2328	chemical diabetes	brittle diabetes	19	194	0.097938
2329	latent diabetes	brittle diabetes	19	194	0.097938
2330	nephrogenic diabetes insipidus	brittle diabetes	19	194	0.097938
2331	bronzed diabetes	brittle diabetes	18	194	0.092784
2332	diabetes	brittle diabetes	18	194	0.092784
2333	glucose	brittle diabetes	18	194	0.092784
2334	diabetes mellitus	necrobiosis lipoidica	19	190	0.1
2335	juvenile diabetes	necrobiosis lipoidica	19	190	0.1
2336	adult-onset diabetes mellitus	necrobiosis lipoidica	19	190	0.1
2337	autoimmune diabetes	necrobiosis lipoidica	19	190	0.1
2338	insulin-dependent diabetes mellitus	necrobiosis lipoidica	19	190	0.1
2339	ketoacidosis-prone diabetes	necrobiosis lipoidica	19	190	0.1
2340	ketoacidosis-resistant diabetes	necrobiosis lipoidica	19	190	0.1
2341	ketoacidosis-resistant diabetes mellitus	necrobiosis lipoidica	19	190	0.1
2342	ketosis-resistant diabetes	necrobiosis lipoidica	19	190	0.1
2343	ketosis-prone diabetes	necrobiosis lipoidica	19	190	0.1
2344	ketosis-resistant diabetes mellitus	necrobiosis lipoidica	19	190	0.1

2345	maturity-onset diabetes mellitus	necrobiosis lipoidica	19	190	0.1
2346	non-insulin-dependent diabetes mellitus	necrobiosis lipoidica	19	190	0.1
2347	diabetes insipidus	necrobiosis lipoidica	19	190	0.1
2348	ketoacidosis prone diabetes	necrobiosis lipoidica	19	190	0.1
2349	sugar diabetes	necrobiosis lipoidica	19	190	0.1
2350	chemical diabetes	necrobiosis lipoidica	19	190	0.1
2351	nephrogenic diabetes insipidus	necrobiosis lipoidica	19	190	0.1
2352	bronzed diabetes	necrobiosis lipoidica	19	190	0.1
2353	diabetes	necrobiosis lipoidica	19	190	0.1
2354	diabetes mellitus	necrobiosis lipoidica diabeticorum	19	190	0.1
2355	juvenile diabetes	necrobiosis lipoidica diabeticorum	19	190	0.1
2356	adult-onset diabetes mellitus	necrobiosis lipoidica diabeticorum	19	190	0.1
2357	autoimmune diabetes	necrobiosis lipoidica diabeticorum	19	190	0.1
2358	insulin-dependent diabetes mellitus	necrobiosis lipoidica diabeticorum	19	190	0.1
2359	ketoacidosis-prone diabetes	necrobiosis lipoidica diabeticorum	19	190	0.1
2360	ketoacidosis-resistant diabetes	necrobiosis lipoidica diabeticorum	19	190	0.1
2361	ketoacidosis-resistant diabetes mellitus	necrobiosis lipoidica diabeticorum	19	190	0.1
2362	ketosis-resistant diabetes	necrobiosis lipoidica diabeticorum	19	190	0.1
2363	ketosis-prone diabetes	necrobiosis lipoidica diabeticorum	19	190	0.1
2364	ketosis-resistant	necrobiosis lipoidica diabeticorum	19	190	0.1

	diabetes mellitus				
2365	maturity-onset diabetes mellitus	necrobiosis lipoidica diabeticorum	19	190	0.1
2366	non-insulin-dependent diabetes mellitus	necrobiosis lipoidica diabeticorum	19	190	0.1
2367	diabetes insipidus	necrobiosis lipoidica diabeticorum	19	190	0.1
2368	ketoacidosis prone diabetes	necrobiosis lipoidica diabeticorum	19	190	0.1
2369	sugar diabetes	necrobiosis lipoidica diabeticorum	19	190	0.1
2370	chemical diabetes	necrobiosis lipoidica diabeticorum	19	190	0.1
2371	nephrogenic diabetes insipidus	necrobiosis lipoidica diabeticorum	19	190	0.1
2372	bronzed diabetes	necrobiosis lipoidica diabeticorum	19	190	0.1
2373	diabetes	necrobiosis lipoidica diabeticorum	19	190	0.1
2374	diabetes mellitus	lypressin	19	187	0.101604
2375	gestational diabetes	lypressin	34	187	0.181818
2376	juvenile diabetes	lypressin	19	187	0.101604
2377	autoimmune diabetes	lypressin	19	187	0.101604
2378	ketoacidosis-prone diabetes	lypressin	19	187	0.101604
2379	ketoacidosis-resistant diabetes	lypressin	19	187	0.101604
2380	ketoacidosis-resistant diabetes mellitus	lypressin	19	187	0.101604
2381	ketosis-resistant diabetes	lypressin	19	187	0.101604
2382	ketosis-prone diabetes	lypressin	19	187	0.101604
2383	ketosis-resistant diabetes mellitus	lypressin	19	187	0.101604

2384	maturity-onset diabetes	lypressin	19	187	0.101604
2385	diabetes insipidus	lypressin	19	187	0.101604
2386	ketoacidosis prone diabetes	lypressin	19	187	0.101604
2387	sugar diabetes	lypressin	19	187	0.101604
2388	chemical diabetes	lypressin	19	187	0.101604
2389	latent diabetes	lypressin	19	187	0.101604
2390	nephrogenic diabetes insipidus	lypressin	19	187	0.101604
2391	bronzed diabetes	lypressin	18	187	0.096257
2392	diabetes	lypressin	18	187	0.096257
2393	diabetes mellitus	hyperglycaemia	19	183	0.103825
2394	gestational diabetes	hyperglycaemia	30	183	0.163934
2395	juvenile diabetes	hyperglycaemia	19	183	0.103825
2396	autoimmune diabetes	hyperglycaemia	19	183	0.103825
2397	ketoacidosis-prone diabetes	hyperglycaemia	19	183	0.103825
2398	ketoacidosis-resistant diabetes	hyperglycaemia	19	183	0.103825
2399	ketoacidosis-resistant diabetes mellitus	hyperglycaemia	19	183	0.103825
2400	ketosis-resistant diabetes	hyperglycaemia	19	183	0.103825
2401	ketosis-prone diabetes	hyperglycaemia	19	183	0.103825
2402	ketosis-resistant diabetes mellitus	hyperglycaemia	19	183	0.103825
2403	diabetes insipidus	hyperglycaemia	19	183	0.103825
2404	ketoacidosis prone diabetes	hyperglycaemia	19	183	0.103825
2405	sugar diabetes	hyperglycaemia	19	183	0.103825
2406	chemical diabetes	hyperglycaemia	19	183	0.103825

2407	nephrogenic diabetes insipidus	hyperglycaemia	19	183	0.103825
2408	bronzed diabetes	hyperglycaemia	18	183	0.098361
2409	glucose tolerance test	hyperglycaemia	17	183	0.092896
2410	diabetes	hyperglycaemia	18	183	0.098361
2411	glucose	hyperglycaemia	17	183	0.092896
2412	diabetes mellitus	metformin	17	182	0.093407
2413	juvenile diabetes	metformin	19	182	0.104396
2414	type i diabetes	metformin	19	182	0.104396
2415	type ii diabetes	metformin	17	182	0.093407
2416	autoimmune diabetes	metformin	19	182	0.104396
2417	ketoacidosis-prone diabetes	metformin	19	182	0.104396
2418	ketoacidosis-resistant diabetes	metformin	19	182	0.104396
2419	ketoacidosis-resistant diabetes mellitus	metformin	17	182	0.093407
2420	ketosis-resistant diabetes	metformin	19	182	0.104396
2421	ketosis-prone diabetes	metformin	19	182	0.104396
2422	ketosis-resistant diabetes mellitus	metformin	17	182	0.093407
2423	diabetes insipidus	metformin	19	182	0.104396
2424	antidiabetic drug	metformin	15	182	0.082418
2425	ketoacidosis prone diabetes	metformin	19	182	0.104396
2426	sugar diabetes	metformin	19	182	0.104396
2427	antidiabetic	metformin	15	182	0.082418
2428	chemical diabetes	metformin	19	182	0.104396
2429	nephrogenic diabetes insipidus	metformin	19	182	0.104396
2430	bronzed diabetes	metformin	19	182	0.104396

2431	diabetes	metformin	19	182	0.104396
2432	diabetes mellitus	liable	16	161	0.099379
2433	juvenile diabetes	liable	18	161	0.111801
2434	autoimmune diabetes	liable	18	161	0.111801
2435	insulin-dependent diabetes mellitus	liable	18	161	0.111801
2436	ketoacidosis-prone diabetes	liable	18	161	0.111801
2437	ketoacidosis-resistant diabetes	liable	18	161	0.111801
2438	ketoacidosis-resistant diabetes mellitus	liable	17	161	0.10559
2439	ketosis-resistant diabetes	liable	18	161	0.111801
2440	ketosis-prone diabetes	liable	18	161	0.111801
2441	ketosis-resistant diabetes mellitus	liable	17	161	0.10559
2442	non-insulin-dependent diabetes	liable	18	161	0.111801
2443	non-insulin-dependent diabetes mellitus	liable	11	161	0.068323
2444	diabetes insipidus	liable	17	161	0.10559
2445	ketoacidosis prone diabetes	liable	17	161	0.10559
2446	sugar diabetes	liable	14	161	0.086957
2447	chemical diabetes	liable	18	161	0.111801
2448	nephrogenic diabetes insipidus	liable	17	161	0.10559
2449	bronzed diabetes	liable	17	161	0.10559
2450	diabetes	liable	17	161	0.10559

2451	diabetes mellitus	phenformin	16	161	0.099379
2452	juvenile diabetes	phenformin	18	161	0.111801
2453	autoimmune diabetes	phenformin	18	161	0.111801
2454	ketoacidosis-prone diabetes	phenformin	18	161	0.111801
2455	ketoacidosis-resistant diabetes	phenformin	18	161	0.111801
2456	ketoacidosis-resistant diabetes mellitus	phenformin	16	161	0.099379
2457	ketosis-resistant diabetes	phenformin	18	161	0.111801
2458	ketosis-prone diabetes	phenformin	18	161	0.111801
2459	ketosis-resistant diabetes mellitus	phenformin	16	161	0.099379
2460	diabetes insipidus	phenformin	17	161	0.10559
2461	ketoacidosis prone diabetes	phenformin	18	161	0.111801
2462	sugar diabetes	phenformin	16	161	0.099379
2463	chemical diabetes	phenformin	18	161	0.111801
2464	nephrogenic diabetes insipidus	phenformin	17	161	0.10559
2465	bronzed diabetes	phenformin	18	161	0.111801
2466	glucose tolerance test	phenformin	12	161	0.074534
2467	insulin shock treatment	phenformin	18	161	0.111801
2468	diabetes	phenformin	18	161	0.111801
2469	glucose	phenformin	14	161	0.086957
2470	diabetes mellitus	glycosuria	17	151	0.112583
2471	gestational diabetes	glycosuria	30	151	0.198676
2472	juvenile diabetes	glycosuria	17	151	0.112583
2473	autoimmune diabetes	glycosuria	17	151	0.112583

2474	ketoacidosis-prone diabetes	glycosuria	17	151	0.112583
2475	ketoacidosis-resistant diabetes	glycosuria	17	151	0.112583
2476	ketoacidosis-resistant diabetes mellitus	glycosuria	17	151	0.112583
2477	ketosis-resistant diabetes	glycosuria	17	151	0.112583
2478	ketosis-prone diabetes	glycosuria	17	151	0.112583
2479	ketosis-resistant diabetes mellitus	glycosuria	17	151	0.112583
2480	diabetes insipidus	glycosuria	17	151	0.112583
2481	ketoacidosis prone diabetes	glycosuria	17	151	0.112583
2482	sugar diabetes	glycosuria	18	151	0.119205
2483	chemical diabetes	glycosuria	17	151	0.112583
2484	nephrogenic diabetes insipidus	glycosuria	17	151	0.112583
2485	glucosuria	glycosuria	1	151	0.006623
2486	bronzed diabetes	glycosuria	16	151	0.10596
2487	diabetes	glycosuria	16	151	0.10596
2488	diabetes mellitus	ketoacidosis	18	149	0.120805
2489	juvenile diabetes	ketoacidosis	17	149	0.114094
2490	adult-onset diabetes mellitus	ketoacidosis	17	149	0.114094
2491	autoimmune diabetes	ketoacidosis	15	149	0.100671
2492	insulin-dependent diabetes mellitus	ketoacidosis	18	149	0.120805
2493	ketosis-resistant diabetes	ketoacidosis	15	149	0.100671
2494	ketosis-prone diabetes	ketoacidosis	15	149	0.100671

2495	ketosis-resistant diabetes mellitus	ketoacidosis	18	149	0.120805
2496	maturity-onset diabetes mellitus	ketoacidosis	17	149	0.114094
2497	non-insulin-dependent diabetes mellitus	ketoacidosis	17	149	0.114094
2498	diabetes insipidus	ketoacidosis	17	149	0.114094
2499	sugar diabetes	ketoacidosis	15	149	0.100671
2500	chemical diabetes	ketoacidosis	15	149	0.100671
2501	nephrogenic diabetes insipidus	ketoacidosis	17	149	0.114094
2502	ketone body	ketoacidosis	24	149	0.161074
2503	diabetic acidosis	ketoacidosis	7	149	0.04698
2504	bronzed diabetes	ketoacidosis	17	149	0.114094
2505	diabetes	ketoacidosis	18	149	0.120805
2506	diabetic diet	carbo loading	16	136	0.117647
2507	carbohydrate loading	carbo loading	16	136	0.117647
2508	low-calorie diet	carbo loading	16	136	0.117647
2509	high-protein diet	carbo loading	16	136	0.117647
2510	low-fat diet	carbo loading	16	136	0.117647
2511	low-salt diet	carbo loading	16	136	0.117647
2512	low-sodium diet	carbo loading	16	136	0.117647
2513	liquid diet	carbo loading	16	136	0.117647
2514	fad diet	carbo loading	16	136	0.117647
2515	obesity diet	carbo loading	16	136	0.117647
2516	bland diet	carbo loading	16	136	0.117647
2517	high-vitamin diet	carbo loading	16	136	0.117647
2518	light diet	carbo loading	16	136	0.117647
2519	allergy diet	carbo loading	16	136	0.117647
2520	macrobiotic diet	carbo loading	16	136	0.117647
2521	ulcer diet	carbo loading	16	136	0.117647
2522	carbohydrate	carbo loading	16	136	0.117647
2523	diabetes mellitus	prediabetes	16	136	0.117647

2524	juvenile diabetes	prediabetes	16	136	0.117647
2525	autoimmune diabetes	prediabetes	16	136	0.117647
2526	growth-onset diabetes	prediabetes	16	136	0.117647
2527	ketoacidosis-prone diabetes	prediabetes	16	136	0.117647
2528	ketoacidosis-resistant diabetes	prediabetes	16	136	0.117647
2529	ketoacidosis-resistant diabetes mellitus	prediabetes	16	136	0.117647
2530	ketosis-resistant diabetes	prediabetes	16	136	0.117647
2531	ketosis-prone diabetes	prediabetes	16	136	0.117647
2532	ketosis-resistant diabetes mellitus	prediabetes	16	136	0.117647
2533	diabetes insipidus	prediabetes	16	136	0.117647
2534	ketoacidosis prone diabetes	prediabetes	16	136	0.117647
2535	sugar diabetes	prediabetes	16	136	0.117647
2536	chemical diabetes	prediabetes	16	136	0.117647
2537	nephrogenic diabetes insipidus	prediabetes	16	136	0.117647
2538	bronzed diabetes	prediabetes	16	136	0.117647
2539	diabetes	prediabetes	16	136	0.117647
2540	diabetic diet	carbohydrate loading	15	117	0.128205
2541	low-calorie diet	carbohydrate loading	15	117	0.128205
2542	high-protein diet	carbohydrate loading	15	117	0.128205
2543	low-fat diet	carbohydrate loading	15	117	0.128205
2544	low-salt diet	carbohydrate loading	14	117	0.119658
2545	low-sodium diet	carbohydrate loading	15	117	0.128205
2546	liquid diet	carbohydrate loading	14	117	0.119658
2547	fad diet	carbohydrate loading	15	117	0.128205
2548	obesity diet	carbohydrate loading	15	117	0.128205
2549	bland diet	carbohydrate loading	15	117	0.128205

2550	high-vitamin diet	carbohydrate loading	15	117	0.128205
2551	light diet	carbohydrate loading	15	117	0.128205
2552	allergy diet	carbohydrate loading	15	117	0.128205
2553	macrobiotic diet	carbohydrate loading	15	117	0.128205
2554	ulcer diet	carbohydrate loading	12	117	0.102564
2555	carbohydrate	carbohydrate loading	14	117	0.119658
2556	diabetic diet	low-calorie diet	15	111	0.135135
2557	high-protein diet	low-calorie diet	14	111	0.126126
2558	low-fat diet	low-calorie diet	13	111	0.117117
2559	low-salt diet	low-calorie diet	14	111	0.126126
2560	low-sodium diet	low-calorie diet	15	111	0.135135
2561	liquid diet	low-calorie diet	13	111	0.117117
2562	fad diet	low-calorie diet	15	111	0.135135
2563	obesity diet	low-calorie diet	15	111	0.135135
2564	reducing diet	low-calorie diet	11	111	0.099099
2565	soft diet	low-calorie diet	12	111	0.108108
2566	bland diet	low-calorie diet	15	111	0.135135
2567	high-vitamin diet	low-calorie diet	15	111	0.135135
2568	light diet	low-calorie diet	15	111	0.135135
2569	allergy diet	low-calorie diet	14	111	0.126126
2570	macrobiotic diet	low-calorie diet	15	111	0.135135
2571	ulcer diet	low-calorie diet	11	111	0.099099
2572	diabetes mellitus	#NAME?	15	125	0.12
2573	juvenile diabetes	#NAME?	16	125	0.128
2574	autoimmune diabetes	#NAME?	15	125	0.12
2575	ketoacidosis-prone diabetes	#NAME?	16	125	0.128
2576	ketoacidosis-resistant diabetes	#NAME?	16	125	0.128
2577	ketoacidosis-resistant diabetes mellitus	#NAME?	15	125	0.12
2578	ketosis-resistant diabetes	#NAME?	15	125	0.12
2579	ketosis-prone diabetes	#NAME?	15	125	0.12

2580	ketosis-resistant diabetes mellitus	#NAME?	15	125	0.12
2581	diabetes insipidus	#NAME?	15	125	0.12
2582	ketoacidosis prone diabetes	#NAME?	16	125	0.128
2583	sugar diabetes	#NAME?	15	125	0.12
2584	chemical diabetes	#NAME?	15	125	0.12
2585	nephrogenic diabetes insipidus	#NAME?	15	125	0.12
2586	bronzed diabetes	#NAME?	15	125	0.12
2587	diabetes	#NAME?	16	125	0.128
2588	acetone	#NAME?	5	125	0.04
2589	diabetes mellitus	cushing's disease	15	124	0.120968
2590	juvenile diabetes	cushing's disease	15	124	0.120968
2591	autoimmune diabetes	cushing's disease	15	124	0.120968
2592	growth-onset diabetes	cushing's disease	16	124	0.129032
2593	ketoacidosis-prone diabetes	cushing's disease	15	124	0.120968
2594	ketoacidosis-resistant diabetes	cushing's disease	15	124	0.120968
2595	ketoacidosis-resistant diabetes mellitus	cushing's disease	15	124	0.120968
2596	ketosis-resistant diabetes	cushing's disease	15	124	0.120968
2597	ketosis-prone diabetes	cushing's disease	15	124	0.120968
2598	ketosis-resistant diabetes mellitus	cushing's disease	15	124	0.120968
2599	diabetes insipidus	cushing's disease	16	124	0.129032
2600	ketoacidosis prone diabetes	cushing's disease	15	124	0.120968
2601	chemical diabetes	cushing's disease	15	124	0.120968

2602	nephrogenic diabetes insipidus	cushing's disease	16	124	0.129032
2603	bronzed diabetes	cushing's disease	15	124	0.120968
2604	obesity diet	cushing's disease	4	124	0.032258
2605	diabetes	cushing's disease	16	124	0.129032
2606	diabetes mellitus	mellitic	15	120	0.125
2607	juvenile diabetes	mellitic	15	120	0.125
2608	autoimmune diabetes	mellitic	15	120	0.125
2609	ketoacidosis-prone diabetes	mellitic	15	120	0.125
2610	ketoacidosis-resistant diabetes	mellitic	15	120	0.125
2611	ketoacidosis-resistant diabetes mellitus	mellitic	15	120	0.125
2612	ketosis-resistant diabetes	mellitic	15	120	0.125
2613	ketosis-prone diabetes	mellitic	15	120	0.125
2614	ketosis-resistant diabetes mellitus	mellitic	15	120	0.125
2615	diabetes insipidus	mellitic	15	120	0.125
2616	ketoacidosis prone diabetes	mellitic	15	120	0.125
2617	sugar diabetes	mellitic	15	120	0.125
2618	chemical diabetes	mellitic	15	120	0.125
2619	nephrogenic diabetes insipidus	mellitic	15	120	0.125
2620	bronzed diabetes	mellitic	15	120	0.125
2621	diabetes	mellitic	15	120	0.125
2622	diabetes mellitus	hypernatremia	15	120	0.125
2623	juvenile diabetes	hypernatremia	15	120	0.125
2624	autoimmune diabetes	hypernatremia	15	120	0.125

2625	ketoacidosis-prone diabetes	hypernatremia	15	120	0.125
2626	ketoacidosis-resistant diabetes	hypernatremia	15	120	0.125
2627	ketoacidosis-resistant diabetes mellitus	hypernatremia	15	120	0.125
2628	ketosis-resistant diabetes	hypernatremia	15	120	0.125
2629	ketosis-prone diabetes	hypernatremia	15	120	0.125
2630	ketosis-resistant diabetes mellitus	hypernatremia	15	120	0.125
2631	diabetes insipidus	hypernatremia	15	120	0.125
2632	ketoacidosis prone diabetes	hypernatremia	15	120	0.125
2633	sugar diabetes	hypernatremia	15	120	0.125
2634	chemical diabetes	hypernatremia	15	120	0.125
2635	nephrogenic diabetes insipidus	hypernatremia	15	120	0.125
2636	bronzed diabetes	hypernatremia	15	120	0.125
2637	diabetes	hypernatremia	15	120	0.125
2638	diabetes mellitus	jambul	15	120	0.125
2639	juvenile diabetes	jambul	15	120	0.125
2640	autoimmune diabetes	jambul	15	120	0.125
2641	ketoacidosis-prone diabetes	jambul	15	120	0.125
2642	ketoacidosis-resistant diabetes	jambul	15	120	0.125
2643	ketoacidosis-resistant diabetes mellitus	jambul	15	120	0.125
2644	ketosis-resistant diabetes	jambul	15	120	0.125

2645	ketosis-prone diabetes	jambul	15	120	0.125
2646	ketosis-resistant diabetes mellitus	jambul	15	120	0.125
2647	diabetes insipidus	jambul	15	120	0.125
2648	ketoacidosis prone diabetes	jambul	15	120	0.125
2649	sugar diabetes	jambul	15	120	0.125
2650	chemical diabetes	jambul	15	120	0.125
2651	nephrogenic diabetes insipidus	jambul	15	120	0.125
2652	bronzed diabetes	jambul	15	120	0.125
2653	diabetes	jambul	15	120	0.125
2654	diabetic diet	kwashiorkor	15	107	0.140187
2655	carbohydrate loading	kwashiorkor	9	107	0.084112
2656	low-calorie diet	kwashiorkor	15	107	0.140187
2657	high-protein diet	kwashiorkor	14	107	0.130841
2658	low-fat diet	kwashiorkor	15	107	0.140187
2659	low-salt diet	kwashiorkor	15	107	0.140187
2660	low-sodium diet	kwashiorkor	15	107	0.140187
2661	fad diet	kwashiorkor	12	107	0.11215
2662	soft diet	kwashiorkor	9	107	0.084112
2663	high-vitamin diet	kwashiorkor	15	107	0.140187
2664	light diet	kwashiorkor	15	107	0.140187
2665	vitamin-deficiency diet	kwashiorkor	13	107	0.121495
2666	dietary	kwashiorkor	12	107	0.11215
2667	macrobiotic diet	kwashiorkor	15	107	0.140187
2668	malnutrition	kwashiorkor	11	107	0.102804
2669	weaning	kwashiorkor	2	107	0.018692
2670	carbohydrate	kwashiorkor	12	107	0.11215
2671	diabetes mellitus	polyuria	14	105	0.133333
2672	juvenile diabetes	polyuria	14	105	0.133333
2673	autoimmune diabetes	polyuria	14	105	0.133333
2674	ketoacidosis-prone diabetes	polyuria	14	105	0.133333

2675	ketoacidosis-resistant diabetes	polyuria	14	105	0.133333
2676	ketoacidosis-resistant diabetes mellitus	polyuria	14	105	0.133333
2677	ketosis-resistant diabetes	polyuria	14	105	0.133333
2678	ketosis-prone diabetes	polyuria	14	105	0.133333
2679	ketosis-resistant diabetes mellitus	polyuria	14	105	0.133333
2680	diabetes insipidus	polyuria	14	105	0.133333
2681	ketoacidosis prone diabetes	polyuria	14	105	0.133333
2682	chemical diabetes	polyuria	14	105	0.133333
2683	nephrogenic diabetes insipidus	polyuria	14	105	0.133333
2684	bronzed diabetes	polyuria	14	105	0.133333
2685	diabetes	polyuria	14	105	0.133333
2686	diabetic diet	high-protein diet	13	89	0.146067
2687	low-calorie diet	high-protein diet	13	89	0.146067
2688	low-fat diet	high-protein diet	13	89	0.146067
2689	low-salt diet	high-protein diet	13	89	0.146067
2690	low-sodium diet	high-protein diet	13	89	0.146067
2691	liquid diet	high-protein diet	12	89	0.134832
2692	fad diet	high-protein diet	13	89	0.146067
2693	obesity diet	high-protein diet	13	89	0.146067
2694	bland diet	high-protein diet	13	89	0.146067
2695	high-vitamin diet	high-protein diet	12	89	0.134832
2696	light diet	high-protein diet	13	89	0.146067
2697	allergy diet	high-protein diet	13	89	0.146067
2698	macrobiotic diet	high-protein diet	13	89	0.146067
2699	malnutrition	high-protein diet	11	89	0.123596
2700	diabetic diet	soft	13	97	0.134021
2701	low-calorie diet	soft	14	97	0.14433
2702	low-fat diet	soft	14	97	0.14433
2703	clear liquid diet	soft	13	97	0.134021

2704	low-salt diet	soft	12	97	0.123711
2705	low-sodium diet	soft	12	97	0.123711
2706	liquid diet	soft	14	97	0.14433
2707	salt-free diet	soft	12	97	0.123711
2708	bland	soft	11	97	0.113402
2709	fad diet	soft	14	97	0.14433
2710	obesity diet	soft	13	97	0.134021
2711	bland diet	soft	11	97	0.113402
2712	light diet	soft	14	97	0.14433
2713	allergy diet	soft	14	97	0.14433
2714	macrobiotic diet	soft	13	97	0.134021
2715	insulin-dependent diabetes mellitus	diabeta	13	91	0.142857
2716	non-insulin-dependent diabetes	diabeta	13	91	0.142857
2717	non-insulin-dependent diabetes mellitus	diabeta	13	91	0.142857
2718	insulin	diabeta	13	91	0.142857
2719	antidiabetic drug	diabeta	13	91	0.142857
2720	lente insulin	diabeta	13	91	0.142857
2721	antidiabetic	diabeta	13	91	0.142857
2722	pancreas	diabeta	13	91	0.142857
2723	insulin reaction	diabeta	13	91	0.142857
2724	glyburide	diabeta	13	91	0.142857
2725	micronase	diabeta	13	91	0.142857
2726	insulin shock	diabeta	25	91	0.274725
2727	recombinant human insulin	diabeta	13	91	0.142857
2728	diabetic diet	low-fat diet	14	84	0.166667
2729	carbohydrate loading	low-fat diet	4	84	0.047619
2730	low-calorie diet	low-fat diet	12	84	0.142857
2731	high-protein diet	low-fat diet	11	84	0.130952
2732	low-salt diet	low-fat diet	13	84	0.154762
2733	low-sodium diet	low-fat diet	13	84	0.154762
2734	liquid diet	low-fat diet	10	84	0.119048
2735	fad diet	low-fat diet	13	84	0.154762
2736	obesity diet	low-fat diet	12	84	0.142857
2737	bland diet	low-fat diet	12	84	0.142857

2738	high-vitamin diet	low-fat diet	12	84	0.142857
2739	light diet	low-fat diet	12	84	0.142857
2740	allergy diet	low-fat diet	12	84	0.142857
2741	macrobiotic diet	low-fat diet	13	84	0.154762
2742	carbohydrate	low-fat diet	5	84	0.059524
2743	insulin-dependent diabetes mellitus	glucagon	12	91	0.131868
2744	non-insulin-dependent diabetes	glucagon	10	91	0.10989
2745	non-insulin-dependent diabetes mellitus	glucagon	10	91	0.10989
2746	insulin	glucagon	11	91	0.120879
2747	lente insulin	glucagon	11	91	0.120879
2748	sugar diabetes	glucagon	13	91	0.142857
2749	pancreas	glucagon	15	91	0.164835
2750	glucose tolerance test	glucagon	6	91	0.065934
2751	insulin reaction	glucagon	11	91	0.120879
2752	islets of langerhans	glucagon	11	91	0.120879
2753	insulin shock	glucagon	24	91	0.263736
2754	islet of langerhans	glucagon	7	91	0.076923
2755	islands of langerhans	glucagon	5	91	0.054945
2756	isles of langerhans	glucagon	7	91	0.076923
2757	recombinant human insulin	glucagon	14	91	0.153846
2758	glucose	glucagon	14	91	0.153846
2759	diabetic diet	pap	15	93	0.16129
2760	low-calorie diet	pap	11	93	0.11828
2761	soft	pap	4	93	0.043011
2762	pablum	pap	8	93	0.086022
2763	liquid diet	pap	14	93	0.150538
2764	bland	pap	7	93	0.075269
2765	fad diet	pap	14	93	0.150538
2766	obesity diet	pap	13	93	0.139785
2767	soft diet	pap	14	93	0.150538
2768	bland diet	pap	14	93	0.150538
2769	light diet	pap	12	93	0.129032

2770	allergy diet	pap	13	93	0.139785
2771	vitamin-deficiency diet	pap	9	93	0.096774
2772	spoon food	pap	11	93	0.11828
2773	macrobiotic diet	pap	15	93	0.16129
2774	ulcer diet	pap	12	93	0.129032
2775	insulin-dependent diabetes mellitus	insulin reaction	13	80	0.1625
2776	diabetic	insulin reaction	5	80	0.0625
2777	insulin	insulin reaction	12	80	0.15
2778	lente insulin	insulin reaction	13	80	0.1625
2779	sugar diabetes	insulin reaction	12	80	0.15
2780	diabetic diet	insulin reaction	6	80	0.075
2781	diabetic coma	insulin reaction	14	80	0.175
2782	diabetic acidosis	insulin reaction	11	80	0.1375
2783	kussmaul's coma	insulin reaction	10	80	0.125
2784	insulin shock	insulin reaction	23	80	0.2875
2785	hypoglycemia	insulin reaction	6	80	0.075
2786	insulin shock therapy	insulin reaction	10	80	0.125
2787	insulin shock treatment	insulin reaction	12	80	0.15
2788	recombinant human insulin	insulin reaction	12	80	0.15
2789	insulin	hypoglycaemia	13	85	0.152941
2790	lente insulin	hypoglycaemia	13	85	0.152941
2791	sugar diabetes	hypoglycaemia	7	85	0.082353
2792	low-calorie diet	hypoglycaemia	12	85	0.141177
2793	high-protein diet	hypoglycaemia	12	85	0.141177
2794	low-fat diet	hypoglycaemia	12	85	0.141177
2795	insulin reaction	hypoglycaemia	13	85	0.152941
2796	insulin shock	hypoglycaemia	25	85	0.294118
2797	low-salt diet	hypoglycaemia	12	85	0.141177
2798	low-sodium diet	hypoglycaemia	12	85	0.141177
2799	soft diet	hypoglycaemia	13	85	0.152941
2800	high-vitamin diet	hypoglycaemia	12	85	0.141177
2801	recombinant human insulin	hypoglycaemia	13	85	0.152941
2802	insulin-dependent	beta cell	13	84	0.154762

	diabetes mellitus				
2803	non-insulin-dependent diabetes	beta cell	13	84	0.154762
2804	non-insulin-dependent diabetes mellitus	beta cell	13	84	0.154762
2805	insulin	beta cell	13	84	0.154762
2806	lente insulin	beta cell	13	84	0.154762
2807	pancreas	beta cell	9	84	0.107143
2808	insulin reaction	beta cell	13	84	0.154762
2809	islets of langerhans	beta cell	11	84	0.130952
2810	insulin shock	beta cell	24	84	0.285714
2811	islet of langerhans	beta cell	10	84	0.119048
2812	islands of langerhans	beta cell	11	84	0.130952
2813	isles of langerhans	beta cell	12	84	0.142857
2814	recombinant human insulin	beta cell	12	84	0.142857
2815	insulin-dependent diabetes mellitus	glyburide	9	72	0.125
2816	non-insulin-dependent diabetes	glyburide	9	72	0.125
2817	non-insulin-dependent diabetes mellitus	glyburide	8	72	0.111111
2818	insulin	glyburide	13	72	0.180556
2819	antidiabetic drug	glyburide	10	72	0.138889
2820	lente insulin	glyburide	13	72	0.180556
2821	antidiabetic	glyburide	10	72	0.138889
2822	pancreas	glyburide	12	72	0.166667
2823	diabeta	glyburide	8	72	0.111111
2824	insulin reaction	glyburide	11	72	0.152778
2825	micronase	glyburide	6	72	0.083333
2826	insulin shock	glyburide	22	72	0.305556
2827	recombinant human insulin	glyburide	12	72	0.166667
2828	insulin-dependent	islets of langerhans	12	77	0.155844

	diabetes mellitus				
2829	non-insulin-dependent diabetes	islets of langerhans	12	77	0.155844
2830	non-insulin-dependent diabetes mellitus	islets of langerhans	12	77	0.155844
2831	insulin	islets of langerhans	12	77	0.155844
2832	lente insulin	islets of langerhans	12	77	0.155844
2833	pancreas	islets of langerhans	11	77	0.142857
2834	glucagon	islets of langerhans	11	77	0.142857
2835	insulin reaction	islets of langerhans	12	77	0.155844
2836	insulin shock	islets of langerhans	23	77	0.298701
2837	insulin shock therapy	islets of langerhans	12	77	0.155844
2838	insulin shock treatment	islets of langerhans	12	77	0.155844
2839	recombinant human insulin	islets of langerhans	12	77	0.155844
2840	insulin-dependent diabetes mellitus	micronase	13	70	0.185714
2841	non-insulin-dependent diabetes	micronase	13	70	0.185714
2842	non-insulin-dependent diabetes mellitus	micronase	13	70	0.185714
2843	insulin	micronase	9	70	0.128571
2844	antidiabetic drug	micronase	8	70	0.114286
2845	lente insulin	micronase	9	70	0.128571
2846	antidiabetic	micronase	8	70	0.114286
2847	pancreas	micronase	9	70	0.128571
2848	diabeta	micronase	5	70	0.071429
2849	insulin reaction	micronase	13	70	0.185714
2850	glyburide	micronase	5	70	0.071429
2851	insulin shock	micronase	23	70	0.328571
2852	recombinant human insulin	micronase	11	70	0.157143
2853	insulin-dependent diabetes mellitus	glipzide	11	71	0.15493

2854	non-insulin-dependent diabetes	glipzide	11	71	0.15493
2855	non-insulin-dependent diabetes mellitus	glipzide	11	71	0.15493
2856	insulin	glipzide	12	71	0.169014
2857	antidiabetic drug	glipzide	12	71	0.169014
2858	lente insulin	glipzide	12	71	0.169014
2859	antidiabetic	glipzide	12	71	0.169014
2860	pancreas	glipzide	12	71	0.169014
2861	insulin reaction	glipzide	11	71	0.15493
2862	insulin shock	glipzide	21	71	0.295775
2863	glucotrol	glipzide	5	71	0.070423
2864	recombinant human insulin	glipzide	11	71	0.15493
2865	diabetic diet	pellagra	14	68	0.205882
2866	low-calorie diet	pellagra	12	68	0.176471
2867	cushing's disease	pellagra	4	68	0.058824
2868	low-fat diet	pellagra	10	68	0.147059
2869	low-salt diet	pellagra	8	68	0.117647
2870	low-sodium diet	pellagra	8	68	0.117647
2871	nicotinic acid	pellagra	8	68	0.117647
2872	fad diet	pellagra	12	68	0.176471
2873	obesity diet	pellagra	7	68	0.102941
2874	light diet	pellagra	11	68	0.161765
2875	allergy diet	pellagra	7	68	0.102941
2876	vitamin-deficiency diet	pellagra	10	68	0.147059
2877	dietary	pellagra	3	68	0.044118
2878	macrobiotic diet	pellagra	13	68	0.191177
2879	malnutrition	pellagra	5	68	0.073529
2880	avitaminosis	pellagra	4	68	0.058824
2881	diabetic diet	reduce	12	68	0.176471
2882	low-calorie diet	reduce	11	68	0.161765
2883	low-fat diet	reduce	12	68	0.176471
2884	clear liquid diet	reduce	10	68	0.147059
2885	liquid diet	reduce	11	68	0.161765
2886	fad diet	reduce	12	68	0.176471
2887	obesity diet	reduce	11	68	0.161765
2888	reducing diet	reduce	10	68	0.147059
2889	bland diet	reduce	9	68	0.132353
2890	light diet	reduce	11	68	0.161765

2891	allergy diet	reduce	10	68	0.147059
2892	macrobiotic diet	reduce	12	68	0.176471
2893	dieting	reduce	5	68	0.073529
2894	insulin-dependent diabetes mellitus	insulin shock	9	45	0.2
2895	insulin	insulin shock	9	45	0.2
2896	lente insulin	insulin shock	9	45	0.2
2897	diabetic coma	insulin shock	9	45	0.2
2898	kussmaul's coma	insulin shock	9	45	0.2
2899	insulin reaction	insulin shock	9	45	0.2
2900	hypoglycemia	insulin shock	9	45	0.2
2901	insulin shock therapy	insulin shock	9	45	0.2
2902	insulin shock treatment	insulin shock	9	45	0.2
2903	recombinant human insulin	insulin shock	9	45	0.2
2904	diabetes mellitus	hand-schuller-christian disease	5	66	0.075758
2905	gestational diabetes	hand-schuller-christian disease	24	66	0.363636
2906	juvenile diabetes	hand-schuller-christian disease	14	66	0.212121
2907	autoimmune diabetes	hand-schuller-christian disease	3	66	0.045455
2908	ketoacidosis-prone diabetes	hand-schuller-christian disease	11	66	0.166667
2909	ketoacidosis-resistant diabetes	hand-schuller-christian disease	11	66	0.166667
2910	ketoacidosis-resistant diabetes mellitus	hand-schuller-christian disease	10	66	0.151515
2911	ketosis-resistant diabetes	hand-schuller-christian disease	3	66	0.045455
2912	ketosis-prone diabetes	hand-schuller-christian disease	3	66	0.045455
2913	diabetes insipidus	hand-schuller-christian disease	10	66	0.151515
2914	ketoacidosis prone diabetes	hand-schuller-christian disease	10	66	0.151515
2915	nephrogenic diabetes insipidus	hand-schuller-christian disease	10	66	0.151515

2916	bronzed diabetes	hand-schuller-christian disease	9	66	0.136364
2917	diabetes	hand-schuller-christian disease	9	66	0.136364
2918	diabetic diet	pablum	11	63	0.174603
2919	pap	pablum	9	63	0.142857
2920	liquid diet	pablum	11	63	0.174603
2921	bland	pablum	11	63	0.174603
2922	fad diet	pablum	11	63	0.174603
2923	obesity diet	pablum	11	63	0.174603
2924	soft diet	pablum	9	63	0.142857
2925	bland diet	pablum	11	63	0.174603
2926	light diet	pablum	11	63	0.174603
2927	allergy diet	pablum	11	63	0.174603
2928	spoon food	pablum	9	63	0.142857
2929	macrobiotic diet	pablum	11	63	0.174603
2930	insulin	hypoglycemia	8	61	0.131148
2931	lente insulin	hypoglycemia	8	61	0.131148
2932	sugar diabetes	hypoglycemia	15	61	0.245902
2933	hypoglycemic agent	hypoglycemia	3	61	0.04918
2934	low-calorie diet	hypoglycemia	7	61	0.114754
2935	high-protein diet	hypoglycemia	6	61	0.098361
2936	low-fat diet	hypoglycemia	8	61	0.131148
2937	insulin reaction	hypoglycemia	8	61	0.131148
2938	insulin shock	hypoglycemia	13	61	0.213115
2939	low-salt diet	hypoglycemia	7	61	0.114754
2940	low-sodium diet	hypoglycemia	6	61	0.098361
2941	soft diet	hypoglycemia	13	61	0.213115
2942	high-vitamin diet	hypoglycemia	7	61	0.114754
2943	recombinant human insulin	hypoglycemia	8	61	0.131148
2944	glucose	hypoglycemia	4	61	0.065574
2945	insulin-dependent diabetes mellitus	glucotrol	11	58	0.189655
2946	non-insulin-dependent diabetes	glucotrol	11	58	0.189655
2947	non-insulin-dependent diabetes mellitus	glucotrol	11	58	0.189655
2948	insulin	glucotrol	11	58	0.189655

2949	antidiabetic drug	glucotrol	7	58	0.12069
2950	lente insulin	glucotrol	11	58	0.189655
2951	antidiabetic	glucotrol	7	58	0.12069
2952	pancreas	glucotrol	11	58	0.189655
2953	insulin reaction	glucotrol	9	58	0.155172
2954	insulin shock	glucotrol	17	58	0.293104
2955	recombinant human insulin	glucotrol	9	58	0.155172
2956	insulin-dependent diabetes mellitus	best	12	63	0.190476
2957	non-insulin-dependent diabetes	best	5	63	0.079365
2958	banting	best	8	63	0.126984
2959	insulin	best	14	63	0.222222
2960	lente insulin	best	14	63	0.222222
2961	insulin reaction	best	10	63	0.15873
2962	insulin shock	best	21	63	0.333333
2963	insulin shock therapy	best	8	63	0.126984
2964	insulin shock treatment	best	8	63	0.126984
2965	macleod	best	3	63	0.047619
2966	john james rickard macleod	best	2	63	0.031746
2967	recombinant human insulin	best	10	63	0.15873
2968	john macleod	best	2	63	0.031746
2969	diabetic diet	clear liquid diet	9	47	0.191489
2970	low-fat diet	clear liquid diet	10	47	0.212766
2971	liquid diet	clear liquid diet	10	47	0.212766
2972	salt-free diet	clear liquid diet	9	47	0.191489
2973	fad diet	clear liquid diet	9	47	0.191489
2974	obesity diet	clear liquid diet	9	47	0.191489
2975	bland diet	clear liquid diet	9	47	0.191489
2976	light diet	clear liquid diet	9	47	0.191489
2977	allergy diet	clear liquid diet	9	47	0.191489
2978	macrobiotic diet	clear liquid diet	9	47	0.191489
2979	ulcer diet	clear liquid diet	2	47	0.042553
2980	diabetic diet	low-salt diet	9	45	0.2
2981	low-sodium diet	low-salt diet	9	45	0.2
2982	liquid diet	low-salt diet	9	45	0.2
2983	salt-free diet	low-salt diet	9	45	0.2

2984	fad diet	low-salt diet	9	45	0.2
2985	obesity diet	low-salt diet	9	45	0.2
2986	bland diet	low-salt diet	9	45	0.2
2987	light diet	low-salt diet	9	45	0.2
2988	allergy diet	low-salt diet	9	45	0.2
2989	macrobiotic diet	low-salt diet	9	45	0.2
2990	diabetic diet	low-sodium diet	9	45	0.2
2991	low-salt diet	low-sodium diet	9	45	0.2
2992	liquid diet	low-sodium diet	9	45	0.2
2993	salt-free diet	low-sodium diet	9	45	0.2
2994	fad diet	low-sodium diet	9	45	0.2
2995	obesity diet	low-sodium diet	9	45	0.2
2996	bland diet	low-sodium diet	9	45	0.2
2997	light diet	low-sodium diet	9	45	0.2
2998	allergy diet	low-sodium diet	9	45	0.2
2999	macrobiotic diet	low-sodium diet	9	45	0.2
3000	insulin-dependent diabetes mellitus	insulin shock therapy	9	45	0.2
3001	insulin	insulin shock therapy	9	45	0.2
3002	lente insulin	insulin shock therapy	9	45	0.2
3003	diabetic coma	insulin shock therapy	9	45	0.2
3004	kussmaul's coma	insulin shock therapy	9	45	0.2
3005	insulin reaction	insulin shock therapy	9	45	0.2
3006	insulin shock	insulin shock therapy	17	45	0.377778
3007	insulin shock treatment	insulin shock therapy	9	45	0.2
3008	recombinant human insulin	insulin shock therapy	9	45	0.2
3009	insulin-dependent diabetes mellitus	sanger	11	54	0.203704
3010	non-insulin-dependent diabetes	sanger	10	54	0.185185
3011	non-insulin-dependent diabetes mellitus	sanger	8	54	0.148148
3012	insulin	sanger	10	54	0.185185
3013	lente insulin	sanger	11	54	0.203704
3014	sir frederick grant banting	sanger	2	54	0.037037

3015	insulin reaction	sanger	10	54	0.185185
3016	insulin shock	sanger	19	54	0.351852
3017	insulin shock therapy	sanger	9	54	0.166667
3018	insulin shock treatment	sanger	8	54	0.148148
3019	recombinant human insulin	sanger	9	54	0.166667
3020	diabetic diet	lite	8	50	0.16
3021	low-calorie diet	lite	10	50	0.2
3022	high-protein diet	lite	10	50	0.2
3023	low-fat diet	lite	10	50	0.2
3024	low-salt diet	lite	10	50	0.2
3025	low-sodium diet	lite	10	50	0.2
3026	soft diet	lite	7	50	0.14
3027	high-vitamin diet	lite	10	50	0.2
3028	light diet	lite	9	50	0.18
3029	light	lite	8	50	0.16
3030	macrobiotic diet	lite	8	50	0.16
3031	diabetic diet	dietetic	10	52	0.192308
3032	liquid diet	dietetic	9	52	0.173077
3033	fad diet	dietetic	10	52	0.192308
3034	obesity diet	dietetic	10	52	0.192308
3035	bland diet	dietetic	9	52	0.173077
3036	light diet	dietetic	10	52	0.192308
3037	dietetical	dietetic	7	52	0.134615
3038	allergy diet	dietetic	10	52	0.192308
3039	dietary	dietetic	10	52	0.192308
3040	macrobiotic diet	dietetic	10	52	0.192308
3041	ulcer diet	dietetic	9	52	0.173077
3042	diabetic diet	nicotinic acid	10	46	0.217391
3043	pellagra	nicotinic acid	3	46	0.065217
3044	liquid diet	nicotinic acid	9	46	0.195652
3045	fad diet	nicotinic acid	10	46	0.217391
3046	obesity diet	nicotinic acid	9	46	0.195652
3047	bland diet	nicotinic acid	7	46	0.152174
3048	high-vitamin diet	nicotinic acid	8	46	0.173913
3049	light diet	nicotinic acid	9	46	0.195652
3050	allergy diet	nicotinic acid	9	46	0.195652
3051	vitamin-deficiency diet	nicotinic acid	8	46	0.173913

3052	macrobiotic diet	nicotinic acid	10	46	0.217391
3053	diabetic diet	liquid diet	9	41	0.219512
3054	low-calorie diet	liquid diet	6	41	0.146342
3055	clear liquid diet	liquid diet	7	41	0.170732
3056	fad diet	liquid diet	9	41	0.219512
3057	obesity diet	liquid diet	9	41	0.219512
3058	bland diet	liquid diet	8	41	0.195122
3059	light diet	liquid diet	9	41	0.219512
3060	allergy diet	liquid diet	9	41	0.219512
3061	macrobiotic diet	liquid diet	9	41	0.219512
3062	ulcer diet	liquid diet	7	41	0.170732
3063	diabetic diet	salt-free diet	9	41	0.219512
3064	low-salt diet	salt-free diet	7	41	0.170732
3065	low-sodium diet	salt-free diet	7	41	0.170732
3066	liquid diet	salt-free diet	7	41	0.170732
3067	fad diet	salt-free diet	9	41	0.219512
3068	obesity diet	salt-free diet	9	41	0.219512
3069	bland diet	salt-free diet	7	41	0.170732
3070	light diet	salt-free diet	9	41	0.219512
3071	allergy diet	salt-free diet	9	41	0.219512
3072	macrobiotic diet	salt-free diet	9	41	0.219512
3073	insulin-dependent diabetes mellitus	islet of langerhans	10	49	0.204082
3074	non-insulin-dependent diabetes	islet of langerhans	9	49	0.183674
3075	non-insulin-dependent diabetes mellitus	islet of langerhans	9	49	0.183674
3076	insulin	islet of langerhans	9	49	0.183674
3077	lente insulin	islet of langerhans	10	49	0.204082
3078	pancreas	islet of langerhans	9	49	0.183674
3079	insulin reaction	islet of langerhans	10	49	0.204082
3080	insulin shock	islet of langerhans	18	49	0.367347
3081	recombinant human insulin	islet of langerhans	9	49	0.183674
3082	secretes	islet of langerhans	4	49	0.081633
3083	diabetes mellitus	ketone+body	12	49	0.244898

3084	adult-onset diabetes mellitus	ketone+body	8	49	0.163265
3085	insulin-dependent diabetes mellitus	ketone+body	9	49	0.183674
3086	ketoacidosis-prone diabetes	ketone+body	1	49	0.020408
3087	ketoacidosis-resistant diabetes	ketone+body	2	49	0.040816
3088	ketoacidosis-resistant diabetes mellitus	ketone+body	12	49	0.244898
3089	ketosis-resistant diabetes mellitus	ketone+body	12	49	0.244898
3090	maturity-onset diabetes mellitus	ketone+body	7	49	0.142857
3091	non-insulin-dependent diabetes mellitus	ketone+body	7	49	0.142857
3092	ketoacidosis prone diabetes	ketone+body	7	49	0.142857
3093	ketone body	ketone+body	7	49	0.142857
3094	acetoacetic acid	ketone+body	6	49	0.122449
3095	beta cell	ketone+body	4	49	0.081633
3096	acetone	ketone+body	4	49	0.081633
3097	diabetic diet	bland	9	44	0.204546
3098	low-fat diet	bland	9	44	0.204546
3099	clear liquid diet	bland	8	44	0.181818
3100	liquid diet	bland	9	44	0.204546
3101	fad diet	bland	9	44	0.204546
3102	obesity diet	bland	9	44	0.204546
3103	soft diet	bland	8	44	0.181818
3104	light diet	bland	9	44	0.204546
3105	allergy diet	bland	9	44	0.204546
3106	macrobiotic diet	bland	9	44	0.204546
3107	diabetic diet	fad diet	8	36	0.222222
3108	liquid diet	fad diet	8	36	0.222222
3109	obesity diet	fad diet	8	36	0.222222
3110	reducing diet	fad diet	8	36	0.222222

3111	bland diet	fad diet	8	36	0.222222
3112	light diet	fad diet	8	36	0.222222
3113	allergy diet	fad diet	8	36	0.222222
3114	macrobiotic diet	fad diet	8	36	0.222222
3115	ulcer diet	fad diet	8	36	0.222222
3116	diabetic diet	obesity diet	8	36	0.222222
3117	low-fat diet	obesity diet	8	36	0.222222
3118	liquid diet	obesity diet	8	36	0.222222
3119	fad diet	obesity diet	8	36	0.222222
3120	bland diet	obesity diet	8	36	0.222222
3121	light diet	obesity diet	8	36	0.222222
3122	allergy diet	obesity diet	8	36	0.222222
3123	macrobiotic diet	obesity diet	8	36	0.222222
3124	ulcer diet	obesity diet	8	36	0.222222
3125	diabetic diet	reducing diet	9	45	0.2
3126	low-fat diet	reducing diet	9	45	0.2
3127	liquid diet	reducing diet	9	45	0.2
3128	fad diet	reducing diet	9	45	0.2
3129	obesity diet	reducing diet	9	45	0.2
3130	bland diet	reducing diet	9	45	0.2
3131	light diet	reducing diet	9	45	0.2
3132	allergy diet	reducing diet	9	45	0.2
3133	macrobiotic diet	reducing diet	9	45	0.2
3134	ulcer diet	reducing diet	9	45	0.2
3135	diabetic diet	soft diet	9	45	0.2
3136	pap	soft diet	9	45	0.2
3137	pablum	soft diet	9	45	0.2
3138	liquid diet	soft diet	9	45	0.2
3139	fad diet	soft diet	9	45	0.2
3140	obesity diet	soft diet	9	45	0.2
3141	bland diet	soft diet	9	45	0.2
3142	light diet	soft diet	9	45	0.2
3143	allergy diet	soft diet	9	45	0.2
3144	macrobiotic diet	soft diet	9	45	0.2
3145	insulin-dependent diabetes mellitus	islands of langerhans	9	45	0.2
3146	non-insulin-dependent diabetes	islands of langerhans	9	45	0.2
3147	non-insulin-dependent	islands of langerhans	9	45	0.2

	diabetes mellitus				
3148	insulin	islands of langerhans	9	45	0.2
3149	lente insulin	islands of langerhans	9	45	0.2
3150	pancreas	islands of langerhans	9	45	0.2
3151	insulin reaction	islands of langerhans	9	45	0.2
3152	insulin shock	islands of langerhans	17	45	0.377778
3153	recombinant human insulin	islands of langerhans	9	45	0.2
3154	insulin-dependent diabetes mellitus	isles of langerhans	9	45	0.2
3155	non-insulin-dependent diabetes	isles of langerhans	9	45	0.2
3156	non-insulin-dependent diabetes mellitus	isles of langerhans	9	45	0.2
3157	insulin	isles of langerhans	9	45	0.2
3158	lente insulin	isles of langerhans	9	45	0.2
3159	pancreas	isles of langerhans	9	45	0.2
3160	insulin reaction	isles of langerhans	9	45	0.2
3161	insulin shock	isles of langerhans	17	45	0.377778
3162	recombinant human insulin	isles of langerhans	9	45	0.2
3163	insulin-dependent diabetes mellitus	hodgkin	9	44	0.204546
3164	non-insulin-dependent diabetes	hodgkin	9	44	0.204546
3165	non-insulin-dependent diabetes mellitus	hodgkin	8	44	0.181818
3166	insulin	hodgkin	9	44	0.204546
3167	lente insulin	hodgkin	9	44	0.204546
3168	chemical diabetes	hodgkin	8	44	0.181818
3169	insulin reaction	hodgkin	9	44	0.204546
3170	insulin shock	hodgkin	17	44	0.386364
3171	recombinant human insulin	hodgkin	9	44	0.204546
3172	diabetic diet	fiber	9	43	0.209302
3173	liquid diet	fiber	8	43	0.186047

3174	fad diet	fiber	9	43	0.209302
3175	obesity diet	fiber	9	43	0.209302
3176	bland diet	fiber	8	43	0.186047
3177	light diet	fiber	9	43	0.209302
3178	allergy diet	fiber	9	43	0.209302
3179	dietary	fiber	9	43	0.209302
3180	roughage	fiber	7	43	0.162791
3181	macrobiotic diet	fiber	9	43	0.209302
3182	diabetic diet	bland diet	8	35	0.228571
3183	liquid diet	bland diet	7	35	0.2
3184	bland	bland diet	7	35	0.2
3185	fad diet	bland diet	8	35	0.228571
3186	obesity diet	bland diet	8	35	0.228571
3187	light diet	bland diet	8	35	0.228571
3188	allergy diet	bland diet	8	35	0.228571
3189	macrobiotic diet	bland diet	8	35	0.228571
3190	ulcer diet	bland diet	8	35	0.228571
3191	diabetic diet	high-vitamin diet	8	35	0.228571
3192	liquid diet	high-vitamin diet	7	35	0.2
3193	fad diet	high-vitamin diet	8	35	0.228571
3194	obesity diet	high-vitamin diet	8	35	0.228571
3195	bland diet	high-vitamin diet	8	35	0.228571
3196	light diet	high-vitamin diet	8	35	0.228571
3197	allergy diet	high-vitamin diet	8	35	0.228571
3198	vitamin-deficiency diet	high-vitamin diet	7	35	0.2
3199	macrobiotic diet	high-vitamin diet	8	35	0.228571
3200	diabetic diet	light diet	8	35	0.228571
3201	high-protein diet	light diet	7	35	0.2
3202	low-fat diet	light diet	8	35	0.228571
3203	fad diet	light diet	8	35	0.228571
3204	obesity diet	light diet	8	35	0.228571
3205	bland diet	light diet	7	35	0.2
3206	high-vitamin diet	light diet	8	35	0.228571
3207	allergy diet	light diet	8	35	0.228571
3208	macrobiotic diet	light diet	8	35	0.228571
3209	insulin	humulin	8	36	0.222222
3210	lente insulin	humulin	8	36	0.222222
3211	diabetic coma	humulin	8	36	0.222222

3212	diabetic acidosis	humulin	8	36	0.222222
3213	insulin reaction	humulin	8	36	0.222222
3214	insulin shock	humulin	15	36	0.416667
3215	insulin shock treatment	humulin	8	36	0.222222
3216	recombinant human insulin	humulin	8	36	0.222222
3217	insulin-dependent diabetes mellitus	insulin shock treatment	1	33	0.030303
3218	insulin	insulin shock treatment	7	33	0.212121
3219	lente insulin	insulin shock treatment	7	33	0.212121
3220	diabetic coma	insulin shock treatment	9	33	0.272727
3221	kussmaul's coma	insulin shock treatment	4	33	0.121212
3222	insulin reaction	insulin shock treatment	7	33	0.212121
3223	insulin shock	insulin shock treatment	15	33	0.454546
3224	insulin shock therapy	insulin shock treatment	8	33	0.242424
3225	recombinant human insulin	insulin shock treatment	7	33	0.212121
3226	diabetic diet	macrobiotic	10	42	0.238095
3227	liquid diet	macrobiotic	8	42	0.190476
3228	fad diet	macrobiotic	10	42	0.238095
3229	obesity diet	macrobiotic	9	42	0.214286
3230	bland diet	macrobiotic	7	42	0.166667
3231	light diet	macrobiotic	9	42	0.214286
3232	allergy diet	macrobiotic	9	42	0.214286
3233	dietary	macrobiotic	6	42	0.142857
3234	ulcer diet	macrobiotic	7	42	0.166667
3235	malnutrition	macrobiotic	7	42	0.166667
3236	macrobiotics	macrobiotic	2	42	0.047619
3237	insulin-dependent diabetes mellitus	macleod	6	39	0.153846
3238	banting	macleod	7	39	0.179487
3239	insulin	macleod	9	39	0.230769
3240	lente insulin	macleod	8	39	0.205128
3241	sir frederick grant banting	macleod	5	39	0.128205
3242	insulin reaction	macleod	9	39	0.230769
3243	insulin shock	macleod	17	39	0.435897
3244	recombinant human insulin	macleod	9	39	0.230769

3245	diabetic diet	dietetical	9	38	0.236842
3246	dietetic	dietetical	2	38	0.052632
3247	liquid diet	dietetical	8	38	0.210526
3248	fad diet	dietetical	8	38	0.210526
3249	obesity diet	dietetical	8	38	0.210526
3250	bland diet	dietetical	8	38	0.210526
3251	light diet	dietetical	8	38	0.210526
3252	allergy diet	dietetical	8	38	0.210526
3253	dietary	dietetical	9	38	0.236842
3254	macrobiotic diet	dietetical	8	38	0.210526
3255	diabetic diet	nonfat	9	40	0.225
3256	low-fat diet	nonfat	6	40	0.15
3257	liquid diet	nonfat	8	40	0.2
3258	salt-free diet	nonfat	6	40	0.15
3259	fad diet	nonfat	9	40	0.225
3260	obesity diet	nonfat	9	40	0.225
3261	bland diet	nonfat	7	40	0.175
3262	light diet	nonfat	8	40	0.2
3263	allergy diet	nonfat	9	40	0.225
3264	macrobiotic diet	nonfat	9	40	0.225
3265	diabetic diet	vegetarianism	8	36	0.222222
3266	liquid diet	vegetarianism	8	36	0.222222
3267	fad diet	vegetarianism	8	36	0.222222
3268	obesity diet	vegetarianism	8	36	0.222222
3269	bland diet	vegetarianism	8	36	0.222222
3270	light diet	vegetarianism	8	36	0.222222
3271	allergy diet	vegetarianism	8	36	0.222222
3272	macrobiotic diet	vegetarianism	8	36	0.222222
3273	ulcer diet	vegetarianism	8	36	0.222222
3274	diabetic diet	allergy diet	7	28	0.25
3275	liquid diet	allergy diet	7	28	0.25
3276	fad diet	allergy diet	7	28	0.25
3277	obesity diet	allergy diet	7	28	0.25
3278	bland diet	allergy diet	7	28	0.25
3279	light diet	allergy diet	7	28	0.25
3280	macrobiotic diet	allergy diet	7	28	0.25
3281	ulcer diet	allergy diet	7	28	0.25
3282	diabetic diet	balanced diet	8	36	0.222222
3283	liquid diet	balanced diet	8	36	0.222222
3284	fad diet	balanced diet	8	36	0.222222
3285	obesity diet	balanced diet	8	36	0.222222

3286	bland diet	balanced diet	8	36	0.222222
3287	light diet	balanced diet	8	36	0.222222
3288	allergy diet	balanced diet	8	36	0.222222
3289	macrobiotic diet	balanced diet	8	36	0.222222
3290	ulcer diet	balanced diet	8	36	0.222222
3291	diabetic diet	vitamin-deficiency diet	7	28	0.25
3292	fad diet	vitamin-deficiency diet	7	28	0.25
3293	obesity diet	vitamin-deficiency diet	7	28	0.25
3294	bland diet	vitamin-deficiency diet	7	28	0.25
3295	high-vitamin diet	vitamin-deficiency diet	7	28	0.25
3296	light diet	vitamin-deficiency diet	7	28	0.25
3297	allergy diet	vitamin-deficiency diet	7	28	0.25
3298	macrobiotic diet	vitamin-deficiency diet	7	28	0.25
3299	diabetic diet	xerophagy	8	36	0.222222
3300	clear liquid diet	xerophagy	8	36	0.222222
3301	liquid diet	xerophagy	8	36	0.222222
3302	fad diet	xerophagy	8	36	0.222222
3303	obesity diet	xerophagy	8	36	0.222222
3304	bland diet	xerophagy	8	36	0.222222
3305	light diet	xerophagy	8	36	0.222222
3306	allergy diet	xerophagy	8	36	0.222222
3307	macrobiotic diet	xerophagy	8	36	0.222222
3308	insulin-dependent diabetes mellitus	fred sanger	8	36	0.222222
3309	non-insulin-dependent diabetes	fred sanger	8	36	0.222222
3310	non-insulin-dependent diabetes mellitus	fred sanger	8	36	0.222222
3311	insulin	fred sanger	8	36	0.222222
3312	lente insulin	fred sanger	8	36	0.222222
3313	insulin reaction	fred sanger	8	36	0.222222
3314	insulin shock	fred sanger	15	36	0.416667
3315	recombinant human insulin	fred sanger	8	36	0.222222
3316	gestational diabetes	schuller-christian disease	16	34	0.470588
3317	juvenile diabetes	schuller-christian disease	10	34	0.294118

3318	ketoacidosis-prone diabetes	schuller-christian disease	6	34	0.176471
3319	ketoacidosis-resistant diabetes	schuller-christian disease	6	34	0.176471
3320	ketoacidosis-resistant diabetes mellitus	schuller-christian disease	5	34	0.147059
3321	diabetes insipidus	schuller-christian disease	6	34	0.176471
3322	ketoacidosis prone diabetes	schuller-christian disease	5	34	0.147059
3323	nephrogenic diabetes insipidus	schuller-christian disease	6	34	0.176471
3324	bronzed diabetes	schuller-christian disease	4	34	0.117647
3325	diabetes	schuller-christian disease	4	34	0.117647
3326	insulin	diabetes	7	28	0.25
3327	lente insulin	diabetes	7	28	0.25
3328	pancreas	diabetes	7	28	0.25
3329	insulin reaction	diabetes	7	28	0.25
3330	insulin shock	diabetes	13	28	0.464286
3331	insulin shock treatment	diabetes	7	28	0.25
3332	recombinant human insulin	diabetes	7	28	0.25
3333	low-calorie diet	low-cal	8	34	0.235294
3334	high-protein diet	low-cal	8	34	0.235294
3335	low-fat diet	low-cal	7	34	0.205882
3336	low-salt diet	low-cal	8	34	0.235294
3337	low-sodium diet	low-cal	8	34	0.235294
3338	lite	low-cal	7	34	0.205882
3339	soft diet	low-cal	7	34	0.205882
3340	high-vitamin diet	low-cal	8	34	0.235294
3341	light	low-cal	7	34	0.205882
3342	diabetic diet	train	9	33	0.272727
3343	high-protein diet	train	6	33	0.181818
3344	clear liquid diet	train	8	33	0.242424
3345	fad diet	train	8	33	0.242424
3346	obesity diet	train	6	33	0.181818
3347	high-vitamin diet	train	6	33	0.181818

3348	light diet	train	8	33	0.242424
3349	allergy diet	train	6	33	0.181818
3350	regimen	train	1	33	0.030303
3351	macrobiotic diet	train	8	33	0.242424
3352	adult-onset diabetes	fire	8	33	0.242424
3353	gestational diabetes	fire	12	33	0.363636
3354	adult-onset diabetes mellitus	fire	5	33	0.151515
3355	growth-onset diabetes	fire	8	33	0.242424
3356	mature-onset diabetes	fire	7	33	0.212121
3357	maturity-onset diabetes	fire	8	33	0.242424
3358	maturity-onset diabetes mellitus	fire	7	33	0.212121
3359	lite	fire	1	33	0.030303
3360	off	fire	5	33	0.151515
3361	light	fire	3	33	0.090909
3362	abstemious	fire	2	33	0.060606
3363	adult-onset diabetes	adolescent	8	33	0.242424
3364	gestational diabetes	adolescent	13	33	0.393939
3365	adult-onset diabetes mellitus	adolescent	8	33	0.242424
3366	growth-onset diabetes	adolescent	7	33	0.212121
3367	mature-onset diabetes	adolescent	8	33	0.242424
3368	maturity-onset diabetes	adolescent	8	33	0.242424
3369	maturity-onset diabetes mellitus	adolescent	8	33	0.242424
3370	juvenile	adolescent	6	33	0.181818
3371	insulin	insulin+shock	7	33	0.212121
3372	lente insulin	insulin+shock	7	33	0.212121
3373	diabetic coma	insulin+shock	7	33	0.212121
3374	kussmaul's coma	insulin+shock	8	33	0.242424
3375	insulin reaction	insulin+shock	8	33	0.242424
3376	insulin shock	insulin+shock	15	33	0.454546

3377	hypoglycemia	insulin+shock	5	33	0.151515
3378	recombinant human insulin	insulin+shock	8	33	0.242424
3379	diabetic diet	dietary	7	30	0.233333
3380	dietetic	dietary	6	30	0.2
3381	fad diet	dietary	7	30	0.233333
3382	obesity diet	dietary	7	30	0.233333
3383	light diet	dietary	8	30	0.266667
3384	dietetical	dietary	1	30	0.033333
3385	allergy diet	dietary	7	30	0.233333
3386	macrobiotic diet	dietary	8	30	0.266667
3387	fare	dietary	4	30	0.133333
3388	dieting	dietary	5	30	0.166667
3389	diabetic diet	arginine	6	32	0.1875
3390	acetoacetic acid	arginine	5	32	0.15625
3391	high-protein diet	arginine	10	32	0.3125
3392	nicotinic acid	arginine	5	32	0.15625
3393	fad diet	arginine	4	32	0.125
3394	light diet	arginine	6	32	0.1875
3395	allergy diet	arginine	1	32	0.03125
3396	macrobiotic diet	arginine	6	32	0.1875
3397	amino acid	arginine	15	32	0.46875
3398	ascorbic acid	arginine	5	32	0.15625
3399	diabetic diet	reichstag	7	31	0.225806
3400	low-calorie diet	reichstag	7	31	0.225806
3401	high-protein diet	reichstag	6	31	0.193548
3402	low-fat diet	reichstag	4	31	0.129032
3403	low-salt diet	reichstag	3	31	0.096774
3404	low-sodium diet	reichstag	7	31	0.225806
3405	fad diet	reichstag	4	31	0.129032
3406	high-vitamin diet	reichstag	8	31	0.258065
3407	light diet	reichstag	8	31	0.258065
3408	macrobiotic diet	reichstag	8	31	0.258065
3409	insulin-dependent diabetes mellitus	frederick sanger	7	29	0.241379
3410	non-insulin-dependent diabetes	frederick sanger	8	29	0.275862

3411	non-insulin-dependent diabetes mellitus	frederick sanger	1	29	0.034483
3412	insulin	frederick sanger	7	29	0.241379
3413	lente insulin	frederick sanger	7	29	0.241379
3414	insulin reaction	frederick sanger	7	29	0.241379
3415	insulin shock	frederick sanger	13	29	0.448276
3416	recombinant human insulin	frederick sanger	7	29	0.241379
3417	low-fat diet	cut	8	29	0.275862
3418	reduce	cut	7	29	0.241379
3419	clear liquid diet	cut	7	29	0.241379
3420	salt-free diet	cut	7	29	0.241379
3421	fad diet	cut	9	29	0.310345
3422	reducing diet	cut	6	29	0.206897
3423	light diet	cut	5	29	0.172414
3424	off	cut	5	29	0.172414
3425	well-fed	cut	3	29	0.103448
3426	macrobiotic diet	cut	1	29	0.034483
3427	banting	john james rickard macleod	7	28	0.25
3428	insulin	john james rickard macleod	7	28	0.25
3429	lente insulin	john james rickard macleod	7	28	0.25
3430	insulin reaction	john james rickard macleod	7	28	0.25
3431	insulin shock	john james rickard macleod	13	28	0.464286
3432	recombinant human insulin	john james rickard macleod	7	28	0.25
3433	banting	c	7	28	0.25
3434	insulin	c	7	28	0.25
3435	lente insulin	c	7	28	0.25
3436	insulin reaction	c	7	28	0.25
3437	insulin shock	c	13	28	0.464286
3438	recombinant human insulin	c	7	28	0.25
3439	banting	h	7	28	0.25
3440	insulin	h	7	28	0.25
3441	lente insulin	h	7	28	0.25
3442	insulin reaction	h	7	28	0.25
3443	insulin shock	h	13	28	0.464286
3444	recombinant human insulin	h	7	28	0.25

3445	banting	best	7	28	0.25
3446	insulin	best	7	28	0.25
3447	lente insulin	best	7	28	0.25
3448	insulin reaction	best	7	28	0.25
3449	insulin shock	best	13	28	0.464286
3450	recombinant human insulin	best	7	28	0.25
3451	diabetic diet	off	8	28	0.285714
3452	low-calorie diet	off	2	28	0.071429
3453	salt-free diet	off	6	28	0.214286
3454	fad diet	off	8	28	0.285714
3455	obesity diet	off	7	28	0.25
3456	reducing diet	off	5	28	0.178571
3457	light diet	off	7	28	0.25
3458	allergy diet	off	6	28	0.214286
3459	macrobiotic diet	off	7	28	0.25
3460	diabetic diet	roughage	7	27	0.259259
3461	low-calorie diet	roughage	7	27	0.259259
3462	fad diet	roughage	7	27	0.259259
3463	obesity diet	roughage	7	27	0.259259
3464	high-vitamin diet	roughage	6	27	0.222222
3465	light diet	roughage	6	27	0.222222
3466	allergy diet	roughage	7	27	0.259259
3467	macrobiotic diet	roughage	7	27	0.259259
3468	diabetic diet	well-fed	6	21	0.285714
3469	low-fat diet	well-fed	6	21	0.285714
3470	fad diet	well-fed	6	21	0.285714
3471	obesity diet	well-fed	6	21	0.285714
3472	light diet	well-fed	6	21	0.285714
3473	allergy diet	well-fed	6	21	0.285714
3474	macrobiotic diet	well-fed	6	21	0.285714
3475	diabetic diet	spoon food	7	25	0.28
3476	pap	spoon food	1	25	0.04
3477	liquid diet	spoon food	3	25	0.12
3478	fad diet	spoon food	7	25	0.28
3479	obesity diet	spoon food	6	25	0.24
3480	bland diet	spoon food	7	25	0.28
3481	light diet	spoon food	6	25	0.24
3482	allergy diet	spoon food	6	25	0.24
3483	macrobiotic diet	spoon food	7	25	0.28
3484	type i diabetes	turn	8	25	0.32

3485	type ii diabetes	turn	6	25	0.24
3486	insulin shock	turn	9	25	0.36
3487	hand-schuller-christian disease	turn	4	25	0.16
3488	insulin shock therapy	turn	6	25	0.24
3489	insulin shock treatment	turn	8	25	0.32
3490	off	turn	3	25	0.12
3491	adult-onset diabetes	stripling	6	24	0.25
3492	gestational diabetes	stripling	11	24	0.458333
3493	adult-onset diabetes mellitus	stripling	6	24	0.25
3494	mature-onset diabetes	stripling	7	24	0.291667
3495	maturity-onset diabetes	stripling	7	24	0.291667
3496	maturity-onset diabetes mellitus	stripling	7	24	0.291667
3497	juvenile	stripling	4	24	0.166667
3498	diabetic diet	light	5	18	0.277778
3499	low-calorie diet	light	6	18	0.333333
3500	low-fat diet	light	5	18	0.277778
3501	lite	light	6	18	0.333333
3502	soft diet	light	6	18	0.333333
3503	abstemious	light	3	18	0.166667
3504	macrobiotic diet	light	5	18	0.277778
3505	diabetic diet	cuttlebone	7	24	0.291667
3506	fad diet	cuttlebone	7	24	0.291667
3507	obesity diet	cuttlebone	6	24	0.25
3508	bland diet	cuttlebone	6	24	0.25
3509	light diet	cuttlebone	6	24	0.25
3510	allergy diet	cuttlebone	6	24	0.25
3511	macrobiotic diet	cuttlebone	7	24	0.291667
3512	supplement	cuttlebone	3	24	0.125
3513	diabetic diet	regimen	7	22	0.318182
3514	fad diet	regimen	6	22	0.272727
3515	obesity diet	regimen	5	22	0.227273
3516	reducing diet	regimen	5	22	0.227273
3517	light diet	regimen	6	22	0.272727
3518	allergy diet	regimen	6	22	0.272727

3519	macrobiotic diet	regimen	7	22	0.318182
3520	beneficial	regimen	2	22	0.090909
3521	diabetic diet	abstemious	7	21	0.333333
3522	clear liquid diet	abstemious	4	21	0.190476
3523	salt-free diet	abstemious	5	21	0.238095
3524	fad diet	abstemious	5	21	0.238095
3525	light diet	abstemious	5	21	0.238095
3526	allergy diet	abstemious	5	21	0.238095
3527	light	abstemious	4	21	0.190476
3528	macrobiotic diet	abstemious	7	21	0.333333
3529	banting	charles herbert best	7	23	0.304348
3530	insulin	charles herbert best	6	23	0.26087
3531	lente insulin	charles herbert best	7	23	0.304348
3532	insulin reaction	charles herbert best	6	23	0.26087
3533	insulin shock	charles herbert best	11	23	0.478261
3534	recombinant human insulin	charles herbert best	6	23	0.26087
3535	insulin	recombinant human insulin	5	16	0.3125
3536	lente insulin	recombinant human insulin	5	16	0.3125
3537	diabetic coma	recombinant human insulin	1	16	0.0625
3538	diabetic acidosis	recombinant human insulin	1	16	0.0625
3539	insulin reaction	recombinant human insulin	5	16	0.3125
3540	insulin shock	recombinant human insulin	9	16	0.5625
3541	insulin shock treatment	recombinant human insulin	5	16	0.3125
3542	diabetic diet	nurture	7	22	0.318182
3543	fad diet	nurture	6	22	0.272727
3544	obesity diet	nurture	6	22	0.272727
3545	bland diet	nurture	4	22	0.181818
3546	light diet	nurture	5	22	0.227273
3547	allergy diet	nurture	6	22	0.272727
3548	train	nurture	3	22	0.136364
3549	macrobiotic diet	nurture	7	22	0.318182
3550	diabetic diet	lowcal	6	21	0.285714
3551	fad diet	lowcal	6	21	0.285714
3552	obesity diet	lowcal	6	21	0.285714
3553	bland diet	lowcal	6	21	0.285714
3554	light diet	lowcal	6	21	0.285714

3555	allergy diet	lowcal	6	21	0.285714
3556	macrobiotic diet	lowcal	6	21	0.285714
3557	diabetic diet	macrobiotic diet	5	15	0.333333
3558	fad diet	macrobiotic diet	5	15	0.333333
3559	obesity diet	macrobiotic diet	5	15	0.333333
3560	bland diet	macrobiotic diet	5	15	0.333333
3561	light diet	macrobiotic diet	5	15	0.333333
3562	allergy diet	macrobiotic diet	5	15	0.333333
3563	diabetic diet	misdiet	6	21	0.285714
3564	fad diet	misdiet	6	21	0.285714
3565	obesity diet	misdiet	6	21	0.285714
3566	bland diet	misdiet	6	21	0.285714
3567	light diet	misdiet	6	21	0.285714
3568	allergy diet	misdiet	6	21	0.285714
3569	macrobiotic diet	misdiet	6	21	0.285714
3570	diabetic diet	nocal	6	21	0.285714
3571	fad diet	nocal	6	21	0.285714
3572	obesity diet	nocal	6	21	0.285714
3573	bland diet	nocal	6	21	0.285714
3574	light diet	nocal	6	21	0.285714
3575	allergy diet	nocal	6	21	0.285714
3576	macrobiotic diet	nocal	6	21	0.285714
3577	diabetic diet	ulcer diet	6	21	0.285714
3578	fad diet	ulcer diet	6	21	0.285714
3579	obesity diet	ulcer diet	6	21	0.285714
3580	bland diet	ulcer diet	6	21	0.285714
3581	light diet	ulcer diet	6	21	0.285714
3582	allergy diet	ulcer diet	6	21	0.285714
3583	macrobiotic diet	ulcer diet	6	21	0.285714
3584	diabetic diet	augsburg	6	21	0.285714
3585	fad diet	augsburg	6	21	0.285714
3586	obesity diet	augsburg	6	21	0.285714
3587	bland diet	augsburg	6	21	0.285714
3588	light diet	augsburg	6	21	0.285714
3589	allergy diet	augsburg	6	21	0.285714
3590	macrobiotic diet	augsburg	6	21	0.285714
3591	diabetic diet	bants	6	21	0.285714
3592	fad diet	bants	6	21	0.285714
3593	obesity diet	bants	6	21	0.285714
3594	bland diet	bants	6	21	0.285714
3595	light diet	bants	6	21	0.285714

3596	allergy diet	bants	6	21	0.285714
3597	macrobiotic diet	bants	6	21	0.285714
3598	diabetic diet	beri	6	21	0.285714
3599	fad diet	beri	6	21	0.285714
3600	obesity diet	beri	6	21	0.285714
3601	bland diet	beri	6	21	0.285714
3602	light diet	beri	6	21	0.285714
3603	allergy diet	beri	6	21	0.285714
3604	macrobiotic diet	beri	6	21	0.285714
3605	diabetic diet	catamount	6	21	0.285714
3606	fad diet	catamount	6	21	0.285714
3607	obesity diet	catamount	6	21	0.285714
3608	bland diet	catamount	6	21	0.285714
3609	light diet	catamount	6	21	0.285714
3610	allergy diet	catamount	6	21	0.285714
3611	macrobiotic diet	catamount	6	21	0.285714
3612	diabetic diet	krill	6	21	0.285714
3613	fad diet	krill	6	21	0.285714
3614	obesity diet	krill	6	21	0.285714
3615	bland diet	krill	6	21	0.285714
3616	light diet	krill	6	21	0.285714
3617	allergy diet	krill	6	21	0.285714
3618	macrobiotic diet	krill	6	21	0.285714
3619	diabetic diet	landtag	6	21	0.285714
3620	fad diet	landtag	6	21	0.285714
3621	obesity diet	landtag	6	21	0.285714
3622	bland diet	landtag	6	21	0.285714
3623	light diet	landtag	6	21	0.285714
3624	allergy diet	landtag	6	21	0.285714
3625	macrobiotic diet	landtag	6	21	0.285714
3626	diabetic diet	dietetics	6	17	0.352941
3627	fad diet	dietetics	5	17	0.294118
3628	obesity diet	dietetics	5	17	0.294118
3629	light diet	dietetics	5	17	0.294118
3630	allergy diet	dietetics	5	17	0.294118
3631	macrobiotic diet	dietetics	6	17	0.352941
3632	sitology	dietetics	2	17	0.117647
3633	diabetic diet	sitology	6	19	0.31579
3634	fad diet	sitology	6	19	0.31579
3635	obesity diet	sitology	5	19	0.263158

3636	light diet	sitology	6	19	0.31579
3637	allergy diet	sitology	5	19	0.263158
3638	macrobiotic diet	sitology	6	19	0.31579
3639	dietetics	sitology	4	19	0.210526
3640	insulin	john macleod	6	20	0.3
3641	lente insulin	john macleod	5	20	0.25
3642	insulin reaction	john macleod	6	20	0.3
3643	insulin shock	john macleod	11	20	0.55
3644	recombinant human insulin	john macleod	6	20	0.3
3645	adult-onset diabetes	strike	6	19	0.31579
3646	adult-onset diabetes mellitus	strike	5	19	0.263158
3647	growth-onset diabetes	strike	6	19	0.31579
3648	mature-onset diabetes	strike	6	19	0.31579
3649	iron-storage disease	strike	4	19	0.210526
3650	hand-schuller-christian disease	strike	1	19	0.052632
3651	lite	strike	5	19	0.263158
3652	abstemious	strike	5	19	0.263158
3653	insulin	produced	5	15	0.333333
3654	lente insulin	produced	5	15	0.333333
3655	insulin reaction	produced	5	15	0.333333
3656	insulin shock	produced	9	15	0.6
3657	recombinant human insulin	produced	5	15	0.333333
3658	ketosis-resistant diabetes	ketone	6	18	0.333333
3659	ketosis-prone diabetes	ketone	6	18	0.333333
3660	ketosis-resistant diabetes mellitus	ketone	6	18	0.333333
3661	chemical diabetes	ketone	6	18	0.333333
3662	ketosis	ketone	4	18	0.222222
3663	ketogenesis	ketone	4	18	0.222222
3664	acetone	ketone	4	18	0.222222
3665	diabetic diet	malnutrition	6	18	0.333333

3666	low-calorie diet	malnutrition	6	18	0.333333
3667	low-fat diet	malnutrition	3	18	0.166667
3668	low-sodium diet	malnutrition	5	18	0.277778
3669	light diet	malnutrition	5	18	0.277778
3670	vitamin-deficiency diet	malnutrition	5	18	0.277778
3671	macrobiotic diet	malnutrition	6	18	0.333333
3672	diabetic diet	weaning	5	14	0.357143
3673	clear liquid diet	weaning	5	14	0.357143
3674	liquid diet	weaning	5	14	0.357143
3675	light diet	weaning	4	14	0.285714
3676	macrobiotic diet	weaning	5	14	0.357143
3677	ablactation	weaning	4	14	0.285714
3678	low-calorie diet	hypoproteinemia	6	18	0.333333
3679	high-protein diet	hypoproteinemia	5	18	0.277778
3680	low-fat diet	hypoproteinemia	6	18	0.333333
3681	low-salt diet	hypoproteinemia	5	18	0.277778
3682	low-sodium diet	hypoproteinemia	6	18	0.333333
3683	soft diet	hypoproteinemia	3	18	0.166667
3684	high-vitamin diet	hypoproteinemia	5	18	0.277778
3685	low-calorie diet	luminescence	6	17	0.352941
3686	low-fat diet	luminescence	6	17	0.352941
3687	low-salt diet	luminescence	5	17	0.294118
3688	low-sodium diet	luminescence	3	17	0.176471
3689	lite	luminescence	5	17	0.294118
3690	light	luminescence	4	17	0.235294
3691	abstemious	luminescence	5	17	0.294118
3692	diabetic diet	spare	6	17	0.352941
3693	fad diet	spare	5	17	0.294118
3694	obesity diet	spare	5	17	0.294118
3695	light diet	spare	5	17	0.294118
3696	allergy diet	spare	5	17	0.294118
3697	macrobiotic diet	spare	6	17	0.352941
3698	avoid	spare	2	17	0.117647
3699	diabetic	xylose	5	15	0.333333
3700	diabetic diet	xylose	5	15	0.333333
3701	diabetic coma	xylose	5	15	0.333333
3702	diabetic acidosis	xylose	5	15	0.333333
3703	carbohydrate loading	xylose	5	15	0.333333

3704	carbohydrate	xylose	5	15	0.333333
3705	insulin	secretes	5	15	0.333333
3706	lente insulin	secretes	5	15	0.333333
3707	insulin reaction	secretes	5	15	0.333333
3708	insulin shock	secretes	9	15	0.6
3709	recombinant human insulin	secretes	5	15	0.333333
3710	adult-onset diabetes	charge	5	16	0.3125
3711	gestational diabetes	charge	10	16	0.625
3712	adult-onset diabetes mellitus	charge	5	16	0.3125
3713	growth-onset diabetes	charge	5	16	0.3125
3714	mature-onset diabetes	charge	5	16	0.3125
3715	non-insulin-dependent diabetes	charge	2	16	0.125
3716	diabetic diet	eat	5	13	0.384615
3717	liquid diet	eat	4	13	0.307692
3718	fad diet	eat	3	13	0.230769
3719	reducing diet	eat	4	13	0.307692
3720	light diet	eat	5	13	0.384615
3721	macrobiotic diet	eat	5	13	0.384615
3722	diabetic diet	nutarian	5	15	0.333333
3723	fad diet	nutarian	5	15	0.333333
3724	obesity diet	nutarian	5	15	0.333333
3725	light diet	nutarian	5	15	0.333333
3726	allergy diet	nutarian	5	15	0.333333
3727	macrobiotic diet	nutarian	5	15	0.333333
3728	diabetic diet	victus	5	15	0.333333
3729	fad diet	victus	5	15	0.333333
3730	obesity diet	victus	5	15	0.333333
3731	light diet	victus	5	15	0.333333
3732	allergy diet	victus	5	15	0.333333
3733	macrobiotic diet	victus	5	15	0.333333
3734	diabetic coma	electroconvulsive therapy	5	15	0.333333
3735	kussmaul's coma	electroconvulsive therapy	5	15	0.333333
3736	insulin shock	electroconvulsive therapy	9	15	0.6

3737	insulin shock therapy	electroconvulsive therapy	5	15	0.333333
3738	insulin shock treatment	electroconvulsive therapy	5	15	0.333333
3739	diabetic coma	ect	5	15	0.333333
3740	kussmaul's coma	ect	5	15	0.333333
3741	insulin shock	ect	9	15	0.6
3742	insulin shock therapy	ect	5	15	0.333333
3743	insulin shock treatment	ect	5	15	0.333333
3744	diabetic coma	electroshock	5	15	0.333333
3745	kussmaul's coma	electroshock	5	15	0.333333
3746	insulin shock	electroshock	9	15	0.6
3747	insulin shock therapy	electroshock	5	15	0.333333
3748	insulin shock treatment	electroshock	5	15	0.333333
3749	diabetic coma	metrazol shock	5	15	0.333333
3750	kussmaul's coma	metrazol shock	5	15	0.333333
3751	insulin shock	metrazol shock	9	15	0.6
3752	insulin shock therapy	metrazol shock	5	15	0.333333
3753	insulin shock treatment	metrazol shock	5	15	0.333333
3754	diabetic coma	metrazol shock therapy	5	15	0.333333
3755	kussmaul's coma	metrazol shock therapy	5	15	0.333333
3756	insulin shock	metrazol shock therapy	9	15	0.6
3757	insulin shock therapy	metrazol shock therapy	5	15	0.333333
3758	insulin shock treatment	metrazol shock therapy	5	15	0.333333
3759	acetoacetic acid	citrulline	5	15	0.333333
3760	nicotinic acid	citrulline	5	15	0.333333
3761	arginine	citrulline	5	15	0.333333
3762	amino acid	citrulline	9	15	0.6
3763	ascorbic acid	citrulline	5	15	0.333333
3764	clear liquid diet	flush	5	15	0.333333
3765	low-salt diet	flush	5	15	0.333333
3766	liquid diet	flush	5	15	0.333333
3767	salt-free diet	flush	5	15	0.333333
3768	bland diet	flush	5	15	0.333333
3769	light diet	flush	5	15	0.333333
3770	diabetic coma	electroshock therapy	4	14	0.285714

3771	kussmaul's coma	electroshock therapy	5	14	0.357143
3772	insulin shock	electroshock therapy	9	14	0.642857
3773	insulin shock therapy	electroshock therapy	4	14	0.285714
3774	insulin shock treatment	electroshock therapy	5	14	0.357143
3775	low-calorie diet	poor	4	14	0.285714
3776	low-fat diet	poor	5	14	0.357143
3777	low-sodium diet	poor	4	14	0.285714
3778	obesity diet	poor	5	14	0.357143
3779	bland diet	poor	5	14	0.357143
3780	light diet	poor	5	14	0.357143
3781	glucose tolerance test	take	4	14	0.285714
3782	hand-schuller-christian disease	take	2	14	0.142857
3783	insulin shock treatment	take	5	14	0.357143
3784	train	take	4	14	0.285714
3785	off	take	6	14	0.428571
3786	eat	take	4	14	0.285714
3787	follow	take	3	14	0.214286
3788	gestational diabetes	teenager	7	12	0.583333
3789	mature-onset diabetes	teenager	4	12	0.333333
3790	maturity-onset diabetes	teenager	5	12	0.416667
3791	maturity-onset diabetes mellitus	teenager	5	12	0.416667
3792	juvenile	teenager	3	12	0.25
3793	low-calorie diet	protein	4	12	0.333333
3794	clear liquid diet	protein	3	12	0.25
3795	low-sodium diet	protein	3	12	0.25
3796	liquid diet	protein	4	12	0.333333
3797	high-vitamin diet	protein	6	12	0.5
3798	amino acid	protein	3	12	0.25
3799	acetoacetic acid	ornithine	5	12	0.416667
3800	nicotinic acid	ornithine	5	12	0.416667
3801	arginine	ornithine	2	12	0.166667
3802	amino acid	ornithine	7	12	0.583333
3803	ascorbic acid	ornithine	4	12	0.333333
3804	iron-storage disease	flat	2	12	0.166667

3805	low-calorie diet	flat	4	12	0.333333
3806	low-salt diet	flat	5	12	0.416667
3807	salt-free diet	flat	5	12	0.416667
3808	bland	flat	4	12	0.333333
3809	bland diet	flat	4	12	0.333333
3810	type i diabetes	lean	2	11	0.181818
3811	low-calorie diet	lean	3	11	0.272727
3812	low-fat diet	lean	5	11	0.454546
3813	clear liquid diet	lean	4	11	0.363636
3814	salt-free diet	lean	4	11	0.363636
3815	spare	lean	4	11	0.363636
3816	diabetic	acetone	3	9	0.333333
3817	diabetic diet	acetone	3	9	0.333333
3818	diabetic coma	acetone	5	9	0.555556
3819	ketone body	acetone	3	9	0.333333
3820	diabetic acidosis	acetone	3	9	0.333333
3821	diabetic diet	supplement	4	10	0.4
3822	high-vitamin diet	supplement	4	10	0.4
3823	light diet	supplement	4	10	0.4
3824	vitamin-deficiency diet	supplement	4	10	0.4
3825	macrobiotic diet	supplement	4	10	0.4
3826	high-protein diet	vitamin k	3	11	0.272727
3827	low-fat diet	vitamin k	4	11	0.363636
3828	clear liquid diet	vitamin k	2	11	0.181818
3829	liquid diet	vitamin k	5	11	0.454546
3830	high-vitamin diet	vitamin k	4	11	0.363636
3831	vitamin-deficiency diet	vitamin k	4	11	0.363636
3832	diabetic diet	amino+acid	4	10	0.4
3833	high-protein diet	amino+acid	4	10	0.4
3834	macrobiotic diet	amino+acid	4	10	0.4
3835	amino acid	amino+acid	7	10	0.7
3836	diabetic diet	macrobiotics	3	9	0.333333
3837	fad diet	macrobiotics	4	9	0.444444
3838	light diet	macrobiotics	4	9	0.444444
3839	dietary	macrobiotics	3	9	0.333333
3840	macrobiotic diet	macrobiotics	4	9	0.444444
3841	acetoacetic acid	cysteine	4	10	0.4

3842	nicotinic acid	cysteine	4	10	0.4
3843	amino acid	cysteine	7	10	0.7
3844	ascorbic acid	cysteine	4	10	0.4
3845	acetoacetic acid	glutamic acid	4	10	0.4
3846	nicotinic acid	glutamic acid	4	10	0.4
3847	amino acid	glutamic acid	7	10	0.7
3848	ascorbic acid	glutamic acid	4	10	0.4
3849	acetoacetic acid	glycine	4	10	0.4
3850	nicotinic acid	glycine	4	10	0.4
3851	amino acid	glycine	7	10	0.7
3852	ascorbic acid	glycine	4	10	0.4
3853	acetoacetic acid	serine	4	10	0.4
3854	nicotinic acid	serine	4	10	0.4
3855	amino acid	serine	7	10	0.7
3856	ascorbic acid	serine	4	10	0.4
3857	acetoacetic acid	thyronine	4	10	0.4
3858	nicotinic acid	thyronine	4	10	0.4
3859	amino acid	thyronine	7	10	0.7
3860	ascorbic acid	thyronine	4	10	0.4
3861	acetoacetic acid	peptide	4	10	0.4
3862	nicotinic acid	peptide	4	10	0.4
3863	amino acid	peptide	7	10	0.7
3864	ascorbic acid	peptide	4	10	0.4
3865	acetoacetic acid	alanine	4	10	0.4
3866	nicotinic acid	alanine	4	10	0.4
3867	amino acid	alanine	7	10	0.7
3868	ascorbic acid	alanine	4	10	0.4
3869	acetoacetic acid	phenylalanine	4	10	0.4
3870	nicotinic acid	phenylalanine	4	10	0.4
3871	amino acid	phenylalanine	7	10	0.7
3872	ascorbic acid	phenylalanine	4	10	0.4
3873	acetoacetic acid	proline	4	10	0.4
3874	nicotinic acid	proline	4	10	0.4
3875	amino acid	proline	7	10	0.7
3876	ascorbic acid	proline	4	10	0.4
3877	acetoacetic acid	valine	4	10	0.4
3878	nicotinic acid	valine	4	10	0.4
3879	amino acid	valine	7	10	0.7
3880	ascorbic acid	valine	4	10	0.4
3881	acetoacetic acid	canavanine	4	10	0.4
3882	nicotinic acid	canavanine	4	10	0.4
3883	amino acid	canavanine	7	10	0.7
3884	ascorbic acid	canavanine	4	10	0.4
3885	acetoacetic acid	cystine	4	10	0.4
3886	nicotinic acid	cystine	4	10	0.4

3887	amino acid	cystine	7	10	0.7
3888	ascorbic acid	cystine	4	10	0.4
3889	acetoacetic acid	aspartic acid	4	10	0.4
3890	nicotinic acid	aspartic acid	4	10	0.4
3891	amino acid	aspartic acid	7	10	0.7
3892	ascorbic acid	aspartic acid	4	10	0.4
3893	acetoacetic acid	histidine	4	10	0.4
3894	nicotinic acid	histidine	4	10	0.4
3895	amino acid	histidine	7	10	0.7
3896	ascorbic acid	histidine	4	10	0.4
3897	acetoacetic acid	hydroxyproline	4	10	0.4
3898	nicotinic acid	hydroxyproline	4	10	0.4
3899	amino acid	hydroxyproline	7	10	0.7
3900	ascorbic acid	hydroxyproline	4	10	0.4
3901	diabetic coma	metrazol shock treatment	5	10	0.5
3902	kussmaul's coma	metrazol shock treatment	3	10	0.3
3903	insulin shock	metrazol shock treatment	4	10	0.4
3904	insulin shock therapy	metrazol shock treatment	4	10	0.4
3905	insulin shock treatment	metrazol shock treatment	4	10	0.4
3906	low-calorie diet	slipper	4	10	0.4
3907	low-fat diet	slipper	4	10	0.4
3908	low-salt diet	slipper	4	10	0.4
3909	low-sodium diet	slipper	4	10	0.4
3910	soft diet	slipper	4	10	0.4
3911	reduce	break	2	10	0.2
3912	insulin shock treatment	break	1	10	0.1
3913	train	break	2	10	0.2
3914	off	break	6	10	0.6
3915	spare	break	4	10	0.4
3916	eat	break	3	10	0.3
3917	avoid	break	2	10	0.2
3918	acetoacetic acid	methionine	4	10	0.4
3919	nicotinic acid	methionine	4	10	0.4
3920	amino acid	methionine	7	10	0.7
3921	ascorbic acid	methionine	4	10	0.4
3922	low-calorie diet	slim	5	9	0.555556
3923	low-fat diet	slim	4	9	0.444444
3924	reduce	slim	3	9	0.333333
3925	low-sodium diet	slim	2	9	0.222222
3926	reducing diet	slim	3	9	0.333333

3927	dieting	slim	1	9	0.111111
3928	carbohydrate loading	food	3	9	0.333333
3929	liquid diet	food	4	9	0.444444
3930	light diet	food	4	9	0.444444
3931	eat	food	3	9	0.333333
3932	fare	food	1	9	0.111111
3933	carbohydrate	food	3	9	0.333333
3934	lite	diet	2	9	0.222222
3935	reichstag	diet	1	9	0.111111
3936	light	diet	1	9	0.111111
3937	regimen	diet	3	9	0.333333
3938	eat	diet	2	9	0.222222
3939	follow	diet	2	9	0.222222
3940	fare	diet	3	9	0.333333
3941	dieted	diet	1	9	0.111111
3942	dieting	diet	1	9	0.111111
3943	legislature	diet	2	9	0.222222
3944	diabetic diet	follow	4	9	0.444444
3945	clear liquid diet	follow	3	9	0.333333
3946	fad diet	follow	3	9	0.333333
3947	light diet	follow	4	9	0.444444
3948	macrobiotic diet	follow	4	9	0.444444
3949	high-protein diet	low	3	9	0.333333
3950	soft	low	3	9	0.333333
3951	bland diet	low	4	9	0.444444
3952	high-vitamin diet	low	4	9	0.444444
3953	light diet	low	4	9	0.444444
3954	diabetic diet	glucose	2	8	0.25
3955	diabetic coma	glucose	3	8	0.375
3956	diabetic acidosis	glucose	3	8	0.375
3957	carbohydrate loading	glucose	4	8	0.5
3958	carbohydrate	glucose	4	8	0.5
3959	adult-onset diabetes	pubescent	4	9	0.444444
3960	gestational diabetes	pubescent	6	9	0.666667
3961	adult-onset diabetes mellitus	pubescent	4	9	0.444444
3962	mature-onset diabetes	pubescent	4	9	0.444444

3963	diabetic diet	ablactation	2	8	0.25
3964	clear liquid diet	ablactation	3	8	0.375
3965	liquid diet	ablactation	2	8	0.25
3966	light diet	ablactation	5	8	0.625
3967	macrobiotic diet	ablactation	3	8	0.375
3968	weaning	ablactation	1	8	0.125
3969	antidiabetic drug	cure	1	8	0.125
3970	cushing's disease	cure	3	8	0.375
3971	hand-schuller-christian disease	cure	3	8	0.375
3972	salt-free diet	cure	1	8	0.125
3973	schuller-christian disease	cure	4	8	0.5
3974	botanical medicine	cure	4	8	0.5
3975	acetoacetic acid	tyrosine	4	8	0.5
3976	nicotinic acid	tyrosine	2	8	0.25
3977	amino acid	tyrosine	5	8	0.625
3978	ascorbic acid	tyrosine	4	8	0.5
3979	acetoacetic acid	dna	4	8	0.5
3980	nicotinic acid	dna	4	8	0.5
3981	amino acid	dna	5	8	0.625
3982	ascorbic acid	dna	2	8	0.25
3983	glucose tolerance test	sugar	3	7	0.428571
3984	carbohydrate loading	sugar	3	7	0.428571
3985	dietary	sugar	1	7	0.142857
3986	glucose	sugar	3	7	0.428571
3987	carbohydrate	sugar	4	7	0.571429
3988	glucose tolerance test	gluconeogenesis	3	7	0.428571
3989	carbohydrate loading	gluconeogenesis	3	7	0.428571
3990	glucose	gluconeogenesis	3	7	0.428571
3991	amino acid	gluconeogenesis	1	7	0.142857
3992	carbohydrate	gluconeogenesis	3	7	0.428571
3993	diabetic diet	amino acid	2	4	0.5
3994	high-protein diet	amino acid	1	4	0.25
3995	light diet	amino acid	3	4	0.75
3996	macrobiotic diet	amino acid	2	4	0.5

3997	ketone body	acetone body	6	7	0.857143
3998	acetoacetic acid	acetone body	1	7	0.142857
3999	beta cell	acetone body	3	7	0.428571
4000	acetone	acetone body	3	7	0.428571
4001	gestational diabetes	youth	5	7	0.714286
4002	maturity-onset diabetes	youth	4	7	0.571429
4003	maturity-onset diabetes mellitus	youth	3	7	0.428571
4004	juvenile	youth	2	7	0.285714
4005	sugar diabetes	sweet	4	7	0.571429
4006	pancreas	sweet	1	7	0.142857
4007	low-fat diet	sweet	3	7	0.428571
4008	low-salt diet	sweet	3	7	0.428571
4009	salt-free diet	sweet	2	7	0.285714
4010	glucose	sweet	1	7	0.142857
4011	diabetic	aleuronat	3	6	0.5
4012	diabetic diet	aleuronat	3	6	0.5
4013	diabetic coma	aleuronat	3	6	0.5
4014	diabetic acidosis	aleuronat	3	6	0.5
4015	diabetic	glycosometer	3	6	0.5
4016	diabetic diet	glycosometer	3	6	0.5
4017	diabetic coma	glycosometer	3	6	0.5
4018	diabetic acidosis	glycosometer	3	6	0.5
4019	adult-onset diabetes	age	3	6	0.5
4020	adult-onset diabetes mellitus	age	3	6	0.5
4021	mature-onset diabetes	age	3	6	0.5
4022	maturity-onset diabetes	age	3	6	0.5
4023	adult-onset diabetes	coup de main	3	6	0.5
4024	adult-onset diabetes mellitus	coup de main	3	6	0.5
4025	growth-onset diabetes	coup de main	3	6	0.5
4026	mature-onset diabetes	coup de main	3	6	0.5
4027	adult-onset diabetes	surprise attack	3	6	0.5

4028	adult-onset diabetes mellitus	surprise attack	3	6	0.5
4029	growth-onset diabetes	surprise attack	3	6	0.5
4030	mature-onset diabetes	surprise attack	3	6	0.5
4031	diabetic diet	fare	3	6	0.5
4032	fad diet	fare	3	6	0.5
4033	light diet	fare	3	6	0.5
4034	macrobiotic diet	fare	3	6	0.5
4035	diabetic diet	ascorbic acid	3	6	0.5
4036	high-vitamin diet	ascorbic acid	3	6	0.5
4037	vitamin-deficiency diet	ascorbic acid	3	6	0.5
4038	macrobiotic diet	ascorbic acid	3	6	0.5
4039	diabetic coma	shock therapy	3	6	0.5
4040	kussmaul's coma	shock therapy	3	6	0.5
4041	insulin shock therapy	shock therapy	3	6	0.5
4042	insulin shock treatment	shock therapy	3	6	0.5
4043	glucose tolerance test	glycogen	3	6	0.5
4044	carbohydrate loading	glycogen	3	6	0.5
4045	glucose	glycogen	3	6	0.5
4046	carbohydrate	glycogen	3	6	0.5
4047	glucose tolerance test	cellulose	3	6	0.5
4048	carbohydrate loading	cellulose	3	6	0.5
4049	glucose	cellulose	3	6	0.5
4050	carbohydrate	cellulose	3	6	0.5
4051	glucose tolerance test	monosaccharide	3	6	0.5
4052	carbohydrate loading	monosaccharide	3	6	0.5
4053	glucose	monosaccharide	3	6	0.5
4054	carbohydrate	monosaccharide	3	6	0.5
4055	glucose tolerance test	invert sugar	3	6	0.5
4056	carbohydrate loading	invert sugar	3	6	0.5
4057	glucose	invert sugar	3	6	0.5

4058	carbohydrate	invert sugar	3	6	0.5
4059	glucose tolerance test	lactose	3	6	0.5
4060	carbohydrate loading	lactose	3	6	0.5
4061	glucose	lactose	3	6	0.5
4062	carbohydrate	lactose	3	6	0.5
4063	acetoacetic acid	acetic	3	6	0.5
4064	clear liquid diet	acetic	3	6	0.5
4065	nicotinic acid	acetic	3	6	0.5
4066	ascorbic acid	acetic	3	6	0.5
4067	acetoacetic acid	base	3	6	0.5
4068	low-salt diet	base	3	6	0.5
4069	nicotinic acid	base	3	6	0.5
4070	ascorbic acid	base	3	6	0.5
4071	high-vitamin diet	beriberi	3	6	0.5
4072	vitamin-deficiency diet	beriberi	3	6	0.5
4073	dietary	beriberi	3	6	0.5
4074	avitaminosis	beriberi	3	6	0.5
4075	carbohydrate loading	conjugated protein	3	6	0.5
4076	amino acid	conjugated protein	5	6	0.833333
4077	carbohydrate	conjugated protein	3	6	0.5
4078	lite	twilight	3	6	0.5
4079	soft diet	twilight	3	6	0.5
4080	light	twilight	3	6	0.5
4081	abstemious	twilight	3	6	0.5
4082	lite	gegenschein	3	6	0.5
4083	soft diet	gegenschein	3	6	0.5
4084	light	gegenschein	3	6	0.5
4085	abstemious	gegenschein	3	6	0.5
4086	lite	blond	3	6	0.5
4087	soft diet	blond	3	6	0.5
4088	light	blond	3	6	0.5
4089	abstemious	blond	3	6	0.5
4090	low-calorie diet	calash	3	6	0.5
4091	low-fat diet	calash	3	6	0.5
4092	low-salt diet	calash	3	6	0.5
4093	low-sodium diet	calash	3	6	0.5
4094	low-calorie diet	victoria	3	6	0.5
4095	low-fat diet	victoria	3	6	0.5
4096	low-salt diet	victoria	3	6	0.5
4097	low-sodium diet	victoria	3	6	0.5
4098	low-calorie diet	dim	3	6	0.5

4099	low-fat diet	dim	3	6	0.5
4100	low-salt diet	dim	3	6	0.5
4101	low-sodium diet	dim	3	6	0.5
4102	low-calorie diet	pale	3	6	0.5
4103	low-fat diet	pale	3	6	0.5
4104	low-salt diet	pale	3	6	0.5
4105	low-sodium diet	pale	3	6	0.5
4106	low-calorie diet	pickup	3	6	0.5
4107	low-fat diet	pickup	3	6	0.5
4108	low-salt diet	pickup	3	6	0.5
4109	low-sodium diet	pickup	3	6	0.5
4110	low-calorie diet	rockaway	3	6	0.5
4111	low-fat diet	rockaway	3	6	0.5
4112	low-salt diet	rockaway	3	6	0.5
4113	low-sodium diet	rockaway	3	6	0.5
4114	low-calorie diet	hertfordshire	3	6	0.5
4115	low-fat diet	hertfordshire	3	6	0.5
4116	low-salt diet	hertfordshire	3	6	0.5
4117	low-sodium diet	hertfordshire	3	6	0.5
4118	low-calorie diet	jones' penstemon	3	6	0.5
4119	low-fat diet	jones' penstemon	3	6	0.5
4120	low-salt diet	jones' penstemon	3	6	0.5
4121	low-sodium diet	jones' penstemon	3	6	0.5
4122	low-calorie diet	penstemon dolius	3	6	0.5
4123	low-fat diet	penstemon dolius	3	6	0.5
4124	low-salt diet	penstemon dolius	3	6	0.5
4125	low-sodium diet	penstemon dolius	3	6	0.5
4126	low-calorie diet	monitor	3	6	0.5
4127	low-fat diet	monitor	3	6	0.5
4128	low-salt diet	monitor	3	6	0.5
4129	low-sodium diet	monitor	3	6	0.5
4130	lite	lamp	3	6	0.5
4131	liquid diet	lamp	3	6	0.5
4132	light	lamp	3	6	0.5
4133	abstemious	lamp	3	6	0.5
4134	clear liquid diet	essential	3	6	0.5
4135	liquid diet	essential	3	6	0.5
4136	amino acid	essential	5	6	0.833333
4137	islets of langerhans	endocrine gland	3	6	0.5
4138	islet of langerhans	endocrine gland	3	6	0.5
4139	islands of langerhans	endocrine gland	3	6	0.5

4140	isles of langerhans	endocrine gland	3	6	0.5
4141	islets of langerhans	island of langerhans	3	6	0.5
4142	islet of langerhans	island of langerhans	3	6	0.5
4143	islands of langerhans	island of langerhans	3	6	0.5
4144	isles of langerhans	island of langerhans	3	6	0.5
4145	pellagra	nicotinamide	3	6	0.5
4146	nicotinic acid	nicotinamide	3	6	0.5
4147	high-vitamin diet	nicotinamide	3	6	0.5
4148	vitamin-deficiency diet	nicotinamide	3	6	0.5
4149	high-vitamin diet	pantothen	3	6	0.5
4150	vitamin-deficiency diet	pantothen	3	6	0.5
4151	amino acid	pantothen	5	6	0.833333
4152	high-protein diet	spartan	2	5	0.4
4153	high-vitamin diet	spartan	3	5	0.6
4154	light diet	spartan	2	5	0.4
4155	abstemious	spartan	3	5	0.6
4156	high-protein diet	tofu	2	5	0.4
4157	low-fat diet	tofu	3	5	0.6
4158	bland	tofu	2	5	0.4
4159	bland diet	tofu	3	5	0.6
4160	clear liquid diet	bright	2	5	0.4
4161	lite	bright	2	5	0.4
4162	light	bright	3	5	0.6
4163	abstemious	bright	3	5	0.6
4164	low-calorie diet	pitch	3	5	0.6
4165	low-fat diet	pitch	3	5	0.6
4166	low-salt diet	pitch	2	5	0.4
4167	high-vitamin diet	pitch	2	5	0.4
4168	soft	gentle	3	5	0.6
4169	bland	gentle	2	5	0.4
4170	soft diet	gentle	2	5	0.4
4171	bland diet	gentle	3	5	0.6
4172	lite	flash	2	5	0.4
4173	fad diet	flash	2	5	0.4
4174	light	flash	3	5	0.6

4175	abstemious	flash	3	5	0.6
4176	low-calorie diet	cool	4	5	0.8
4177	reduce	cool	1	5	0.2
4178	low-salt diet	cool	2	5	0.4
4179	salt-free diet	cool	2	5	0.4
4180	off	cool	1	5	0.2
4181	low-calorie diet	macerate	3	5	0.6
4182	low-sodium diet	macerate	2	5	0.4
4183	liquid diet	macerate	3	5	0.6
4184	light diet	macerate	2	5	0.4
4185	low-calorie diet	nyctalopia	3	5	0.6
4186	low-sodium diet	nyctalopia	2	5	0.4
4187	high-vitamin diet	nyctalopia	3	5	0.6
4188	vitamin-deficiency diet	nyctalopia	2	5	0.4
4189	acetoacetic acid	bitartrate	3	5	0.6
4190	low-salt diet	bitartrate	2	5	0.4
4191	nicotinic acid	bitartrate	3	5	0.6
4192	ascorbic acid	bitartrate	2	5	0.4
4193	acetoacetic acid	saponify	3	5	0.6
4194	low-salt diet	saponify	2	5	0.4
4195	nicotinic acid	saponify	2	5	0.4
4196	ascorbic acid	saponify	3	5	0.6
4197	diabetic diet	avoid	2	5	0.4
4198	light diet	avoid	2	5	0.4
4199	balanced diet	avoid	3	5	0.6
4200	macrobiotic diet	avoid	3	5	0.6
4201	diabetic diet	protestant	2	5	0.4
4202	salt-free diet	protestant	2	5	0.4
4203	light diet	protestant	3	5	0.6
4204	macrobiotic diet	protestant	3	5	0.6
4205	diabetic diet	dieted	2	5	0.4
4206	light diet	dieted	3	5	0.6
4207	regimen	dieted	2	5	0.4
4208	macrobiotic diet	dieted	3	5	0.6
4209	non-insulin-dependent diabetes	positive	1	5	0.2
4210	non-insulin-dependent diabetes mellitus	positive	1	5	0.2
4211	lite	positive	1	5	0.2

4212	light diet	positive	3	5	0.6
4213	vitamin-deficiency diet	positive	1	5	0.2
4214	regimen	positive	1	5	0.2
4215	abstemious	positive	1	5	0.2
4216	beneficial	positive	1	5	0.2
4217	non-insulin-dependent diabetes	form	3	5	0.6
4218	ketone body	form	3	5	0.6
4219	iron-storage disease	form	1	5	0.2
4220	train	form	2	5	0.4
4221	maturity-onset diabetes	rear	2	4	0.5
4222	maturity-onset diabetes mellitus	rear	2	4	0.5
4223	fred sanger	rear	1	4	0.25
4224	nurture	rear	3	4	0.75
4225	maturity-onset diabetes	force	2	4	0.5
4226	ketone body	force	3	4	0.75
4227	salt-free diet	force	1	4	0.25
4228	fad diet	force	1	4	0.25
4229	pancreas	sweetbread	3	4	0.75
4230	islets of langerhans	sweetbread	1	4	0.25
4231	recombinant human insulin	sweetbread	2	4	0.5
4232	secretes	sweetbread	2	4	0.5
4233	adult-onset diabetes	blitz	3	4	0.75
4234	adult-onset diabetes mellitus	blitz	1	4	0.25
4235	growth-onset diabetes	blitz	2	4	0.5
4236	mature-onset diabetes	blitz	2	4	0.5
4237	adult-onset diabetes	adolescence	1	4	0.25
4238	growth-onset diabetes	adolescence	3	4	0.75
4239	maturity-onset diabetes	adolescence	2	4	0.5
4240	maturity-onset diabetes mellitus	adolescence	2	4	0.5

4241	adult-onset diabetes	access	2	4	0.5
4242	growth-onset diabetes	access	2	4	0.5
4243	mature-onset diabetes	access	3	4	0.75
4244	iron-storage disease	access	1	4	0.25
4245	adult-onset diabetes	down	1	4	0.25
4246	growth-onset diabetes	down	1	4	0.25
4247	soft	down	2	4	0.5
4248	off	down	2	4	0.5
4249	eat	down	2	4	0.5
4250	type i diabetes	foot	3	4	0.75
4251	type ii diabetes	foot	1	4	0.25
4252	best	foot	2	4	0.5
4253	diabetic diet	avitaminosis	1	4	0.25
4254	high-vitamin diet	avitaminosis	2	4	0.5
4255	vitamin-deficiency diet	avitaminosis	2	4	0.5
4256	macrobiotic diet	avitaminosis	3	4	0.75
4257	diabetic diet	ants	2	4	0.5
4258	low-calorie diet	ants	3	4	0.75
4259	low-sodium diet	ants	1	4	0.25
4260	macrobiotic diet	ants	2	4	0.5
4261	ketone body	sea	3	4	0.75
4262	low-salt diet	sea	1	4	0.25
4263	salt-free diet	sea	3	4	0.75
4264	glucose tolerance test	starch	1	4	0.25
4265	carbohydrate loading	starch	2	4	0.5
4266	glucose	starch	3	4	0.75
4267	carbohydrate	starch	2	4	0.5
4268	hypoglycaemic agent	preventive	2	4	0.5
4269	hypoglycemic agent	preventive	2	4	0.5
4270	cushing's disease	preventive	1	4	0.25
4271	botanical medicine	preventive	3	4	0.75
4272	diabetic coma	shock+therapy	3	4	0.75

4273	kussmaul's coma	shock+therapy	1	4	0.25
4274	insulin shock therapy	shock+therapy	2	4	0.5
4275	insulin shock treatment	shock+therapy	2	4	0.5
4276	high-vitamin diet	choline	2	4	0.5
4277	vitamin-deficiency diet	choline	3	4	0.75
4278	macrobiotic diet	choline	2	4	0.5
4279	supplement	choline	1	4	0.25
4280	salt-free diet	fall	2	4	0.5
4281	off	fall	3	4	0.75
4282	spoon food	fall	1	4	0.25
4283	light	fall	2	4	0.5
4284	low-calorie diet	spar	3	4	0.75
4285	low-fat diet	spar	2	4	0.5
4286	low-sodium diet	spar	2	4	0.5
4287	athletic training	spar	1	4	0.25
4288	carbohydrate loading	riboflavin	3	4	0.75
4289	high-vitamin diet	riboflavin	2	4	0.5
4290	vitamin-deficiency diet	riboflavin	2	4	0.5
4291	carbohydrate	riboflavin	1	4	0.25
4292	acetoacetic acid	asparagine	1	4	0.25
4293	amino acid	asparagine	3	4	0.75
4294	ascorbic acid	asparagine	3	4	0.75
4295	best	fat	1	4	0.25
4296	obesity diet	fat	2	4	0.5
4297	dietary	fat	1	4	0.25
4298	well-fed	fat	2	4	0.5
4299	insulin shock	galvanic	3	4	0.75
4300	insulin shock treatment	galvanic	3	4	0.75
4301	produced	galvanic	1	4	0.25
4302	high-protein diet	glow	1	4	0.25
4303	lite	glow	2	4	0.5
4304	high-vitamin diet	glow	1	4	0.25
4305	light	glow	2	4	0.5
4306	abstemious	glow	2	4	0.5
4307	lite	shade	2	3	0.666667

4308	light	shade	2	3	0.666667
4309	abstemious	shade	2	3	0.666667
4310	lite	shine	2	3	0.666667
4311	light	shine	2	3	0.666667
4312	abstemious	shine	2	3	0.666667
4313	lite	beam	2	3	0.666667
4314	light	beam	2	3	0.666667
4315	abstemious	beam	2	3	0.666667
4316	lite	chiaroscuro	2	3	0.666667
4317	light	chiaroscuro	2	3	0.666667
4318	abstemious	chiaroscuro	2	3	0.666667
4319	lite	counterglow	2	3	0.666667
4320	light	counterglow	2	3	0.666667
4321	abstemious	counterglow	2	3	0.666667
4322	lite	dawn	2	3	0.666667
4323	light	dawn	2	3	0.666667
4324	abstemious	dawn	2	3	0.666667
4325	lite	firelight	2	3	0.666667
4326	light	firelight	2	3	0.666667
4327	abstemious	firelight	2	3	0.666667
4328	lite	flare	2	3	0.666667
4329	light	flare	2	3	0.666667
4330	abstemious	flare	2	3	0.666667
4331	lite	gleam	2	3	0.666667
4332	light	gleam	2	3	0.666667
4333	abstemious	gleam	2	3	0.666667
4334	lite	halo	2	3	0.666667
4335	light	halo	2	3	0.666667
4336	abstemious	halo	2	3	0.666667
4337	lite	illuminate	2	3	0.666667
4338	light	illuminate	2	3	0.666667
4339	abstemious	illuminate	2	3	0.666667
4340	lite	lens	2	3	0.666667
4341	light	lens	2	3	0.666667
4342	abstemious	lens	2	3	0.666667
4343	lite	ray	2	3	0.666667
4344	light	ray	2	3	0.666667
4345	abstemious	ray	2	3	0.666667
4346	lite	refraction	2	3	0.666667
4347	light	refraction	2	3	0.666667
4348	abstemious	refraction	2	3	0.666667
4349	lite	shaft	2	3	0.666667
4350	light	shaft	2	3	0.666667
4351	abstemious	shaft	2	3	0.666667
4352	lite	starlight	2	3	0.666667

4353	light	starlight	2	3	0.666667
4354	abstemious	starlight	2	3	0.666667
4355	lite	sun	2	3	0.666667
4356	light	sun	2	3	0.666667
4357	abstemious	sun	2	3	0.666667
4358	insulin shock	electrocute	3	3	1
4359	insulin shock treatment	electrocute	2	3	0.666667
4360	insulin shock	air spring	3	3	1
4361	insulin shock treatment	air spring	2	3	0.666667
4362	insulin shock	electrify	3	3	1
4363	insulin shock treatment	electrify	2	3	0.666667
4364	insulin shock	sonic boom	3	3	1
4365	insulin shock treatment	sonic boom	2	3	0.666667
4366	insulin shock	air cushion	3	3	1
4367	insulin shock treatment	air cushion	2	3	0.666667
4368	insulin shock	revolt	3	3	1
4369	insulin shock treatment	revolt	2	3	0.666667
4370	insulin shock	shockproof	3	3	1
4371	insulin shock treatment	shockproof	2	3	0.666667
4372	insulin shock	bump	3	3	1
4373	insulin shock treatment	bump	2	3	0.666667
4374	insulin shock	galvanize	3	3	1
4375	insulin shock treatment	galvanize	2	3	0.666667
4376	insulin shock	stagger	3	3	1
4377	insulin shock treatment	stagger	2	3	0.666667
4378	insulin shock	jar	3	3	1
4379	insulin shock treatment	jar	2	3	0.666667
4380	insulin shock	dashpot	3	3	1
4381	insulin shock treatment	dashpot	2	3	0.666667
4382	salt-free diet	slender	2	3	0.666667
4383	light diet	slender	2	3	0.666667
4384	abstemious	slender	2	3	0.666667
4385	clear liquid diet	blank	2	3	0.666667
4386	bland	blank	2	3	0.666667
4387	bland diet	blank	2	3	0.666667

4388	high-protein diet	bude light	2	3	0.666667
4389	low-fat diet	bude light	2	3	0.666667
4390	high-vitamin diet	bude light	2	3	0.666667
4391	low-salt diet	flux	2	3	0.666667
4392	liquid diet	flux	2	3	0.666667
4393	salt-free diet	flux	2	3	0.666667
4394	pellagra	deficiency disease	2	3	0.666667
4395	high-vitamin diet	deficiency disease	2	3	0.666667
4396	vitamin-deficiency diet	deficiency disease	2	3	0.666667
4397	low-fat diet	owe	2	3	0.666667
4398	clear liquid diet	owe	2	3	0.666667
4399	liquid diet	owe	2	3	0.666667
4400	low-fat diet	sago	2	3	0.666667
4401	clear liquid diet	sago	2	3	0.666667
4402	liquid diet	sago	2	3	0.666667
4403	salt-free diet	smooth	2	3	0.666667
4404	bland	smooth	2	3	0.666667
4405	bland diet	smooth	2	3	0.666667
4406	soft	mild	2	3	0.666667
4407	bland	mild	2	3	0.666667
4408	bland diet	mild	2	3	0.666667
4409	insulin shock	dumb	3	3	1
4410	insulin shock treatment	dumb	2	3	0.666667
4411	insulin shock	recover	3	3	1
4412	insulin shock treatment	recover	2	3	0.666667
4413	high-protein diet	ting	2	3	0.666667
4414	high-vitamin diet	ting	2	3	0.666667
4415	light diet	ting	2	3	0.666667
4416	vitamin-deficiency diet	folacin	2	3	0.666667
4417	amino acid	folacin	3	3	1
4418	low-sodium diet	hypocalcaemia	2	3	0.666667
4419	high-vitamin diet	hypocalcaemia	2	3	0.666667
4420	vitamin-deficiency diet	hypocalcaemia	2	3	0.666667
4421	low-sodium diet	hypocalcemia	2	3	0.666667
4422	high-vitamin diet	hypocalcemia	2	3	0.666667

4423	vitamin-deficiency diet	hypocalcemia	2	3	0.666667
4424	vitamin-deficiency diet	rickets	2	3	0.666667
4425	malnutrition	rickets	2	3	0.666667
4426	avitaminosis	rickets	2	3	0.666667
4427	light diet	starve	2	3	0.666667
4428	vitamin-deficiency diet	starve	2	3	0.666667
4429	malnutrition	starve	2	3	0.666667
4430	nicotinic acid	tryptophan	2	3	0.666667
4431	amino acid	tryptophan	3	3	1
4432	nicotinic acid	tryptophane	2	3	0.666667
4433	amino acid	tryptophane	3	3	1
4434	best	first class	3	3	1
4435	train	first class	1	3	0.333333
4436	charles herbert best	first class	1	3	0.333333
4437	best	pride	2	3	0.666667
4438	charles herbert best	pride	2	3	0.666667
4439	light diet	calcium	2	3	0.666667
4440	dietary	calcium	2	3	0.666667
4441	supplement	calcium	2	3	0.666667
4442	low-salt diet	iron	2	3	0.666667
4443	salt-free diet	iron	1	3	0.333333
4444	dietary	iron	2	3	0.666667
4445	supplement	iron	1	3	0.333333
4446	dietary	leucine	2	3	0.666667
4447	amino acid	leucine	3	3	1
4448	high-vitamin diet	rda	2	3	0.666667
4449	vitamin-deficiency diet	rda	2	3	0.666667
4450	dietary	rda	2	3	0.666667
4451	soft	mush	2	3	0.666667
4452	pap	mush	2	3	0.666667
4453	pablum	mush	2	3	0.666667
4454	carbohydrate loading	sucrose	2	3	0.666667
4455	glucose	sucrose	2	3	0.666667
4456	carbohydrate	sucrose	2	3	0.666667
4457	carbohydrate loading	seminose	2	3	0.666667
4458	glucose	seminose	2	3	0.666667
4459	carbohydrate	seminose	2	3	0.666667

4460	carbohydrate loading	beet sugar	2	3	0.666667
4461	glucose	beet sugar	2	3	0.666667
4462	carbohydrate	beet sugar	2	3	0.666667
4463	carbohydrate loading	thiamine	2	3	0.666667
4464	vitamin-deficiency diet	thiamine	2	3	0.666667
4465	carbohydrate	thiamine	2	3	0.666667
4466	low-calorie diet	idiot light	2	3	0.666667
4467	low-fat diet	idiot light	2	3	0.666667
4468	low-sodium diet	idiot light	2	3	0.666667
4469	low-calorie diet	idiot+light	2	3	0.666667
4470	low-fat diet	idiot+light	2	3	0.666667
4471	low-sodium diet	idiot+light	2	3	0.666667
4472	lite	candlelight	2	3	0.666667
4473	soft diet	candlelight	2	3	0.666667
4474	abstemious	candlelight	2	3	0.666667
4475	low-calorie diet	evil	3	3	1
4476	low-sodium diet	evil	1	3	0.333333
4477	high-vitamin diet	evil	1	3	0.333333
4478	ulcer diet	evil	1	3	0.333333
4479	lite	dark	2	3	0.666667
4480	light	dark	2	3	0.666667
4481	abstemious	dark	2	3	0.666667
4482	low-calorie diet	faint	2	3	0.666667
4483	low-sodium diet	faint	2	3	0.666667
4484	light	faint	2	3	0.666667
4485	low-calorie diet	pump	2	3	0.666667
4486	low-sodium diet	pump	2	3	0.666667
4487	reducing diet	pump	2	3	0.666667
4488	high-vitamin diet	scurvy	1	3	0.333333
4489	vitamin-deficiency diet	scurvy	3	3	1
4490	ascorbic acid	scurvy	1	3	0.333333
4491	avitaminosis	scurvy	1	3	0.333333
4492	low-calorie diet	simple	2	3	0.666667
4493	salt-free diet	simple	2	3	0.666667
4494	light diet	simple	2	3	0.666667
4495	glucose tolerance test	narrow	2	3	0.666667
4496	carbohydrate loading	narrow	2	3	0.666667
4497	carbohydrate	narrow	2	3	0.666667

4498	soft	mellow	2	3	0.666667
4499	salt-free diet	mellow	2	3	0.666667
4500	soft diet	mellow	2	3	0.666667
4501	hypoglycaemic agent	yeast	2	3	0.666667
4502	hypoglycemic agent	yeast	2	3	0.666667
4503	dietary	yeast	2	3	0.666667
4504	hypoglycaemic agent	emetic	2	3	0.666667
4505	hypoglycemic agent	emetic	2	3	0.666667
4506	botanical medicine	emetic	2	3	0.666667
4507	beta cell	radiotherapy	1	3	0.333333
4508	hand-schuller-christian disease	radiotherapy	3	3	1
4509	insulin shock therapy	radiotherapy	1	3	0.333333
4510	schuller-christian disease	radiotherapy	1	3	0.333333
4511	hypoglycaemic agent	principal	2	3	0.666667
4512	hypoglycemic agent	principal	2	3	0.666667
4513	liable	principal	2	3	0.666667
4514	hypoglycaemic agent	cathartic	2	3	0.666667
4515	hypoglycemic agent	cathartic	2	3	0.666667
4516	botanical medicine	cathartic	2	3	0.666667
4517	hypoglycaemic agent	dose	2	3	0.666667
4518	hypoglycemic agent	dose	2	3	0.666667
4519	botanical medicine	dose	2	3	0.666667
4520	acetoacetic acid	acetic acid	2	3	0.666667
4521	nicotinic acid	acetic acid	2	3	0.666667
4522	ascorbic acid	acetic acid	2	3	0.666667
4523	acetoacetic acid	sour	2	3	0.666667
4524	nicotinic acid	sour	2	3	0.666667
4525	ascorbic acid	sour	2	3	0.666667
4526	hypoglycaemic agent	pathogen	2	3	0.666667

4527	hypoglycemic agent	pathogen	2	3	0.666667
4528	cushing's disease	pathogen	2	3	0.666667
4529	acetoacetic acid	stearic acid	2	3	0.666667
4530	nicotinic acid	stearic acid	2	3	0.666667
4531	ascorbic acid	stearic acid	2	3	0.666667
4532	acetoacetic acid	subacid	2	3	0.666667
4533	nicotinic acid	subacid	2	3	0.666667
4534	ascorbic acid	subacid	2	3	0.666667
4535	acetoacetic acid	succinic acid	2	3	0.666667
4536	nicotinic acid	succinic acid	2	3	0.666667
4537	ascorbic acid	succinic acid	2	3	0.666667
4538	acetoacetic acid	tartrate	2	3	0.666667
4539	nicotinic acid	tartrate	2	3	0.666667
4540	ascorbic acid	tartrate	2	3	0.666667
4541	acetoacetic acid	urate	2	3	0.666667
4542	nicotinic acid	urate	2	3	0.666667
4543	ascorbic acid	urate	2	3	0.666667
4544	acetoacetic acid	#NAME?	2	3	0.666667
4545	nicotinic acid	#NAME?	2	3	0.666667
4546	ascorbic acid	#NAME?	2	3	0.666667
4547	acetoacetic acid	acid-forming	2	3	0.666667
4548	nicotinic acid	acid-forming	2	3	0.666667
4549	ascorbic acid	acid-forming	2	3	0.666667
4550	acetoacetic acid	manganic acid	2	3	0.666667
4551	nicotinic acid	manganic acid	2	3	0.666667
4552	ascorbic acid	manganic acid	2	3	0.666667
4553	acetoacetic acid	margaric acid	2	3	0.666667
4554	nicotinic acid	margaric acid	2	3	0.666667
4555	ascorbic acid	margaric acid	2	3	0.666667
4556	acetoacetic acid	mucic acid	2	3	0.666667
4557	nicotinic acid	mucic acid	2	3	0.666667
4558	ascorbic acid	mucic acid	2	3	0.666667
4559	acetoacetic acid	nitric acid	2	3	0.666667
4560	nicotinic acid	nitric acid	2	3	0.666667
4561	ascorbic acid	nitric acid	2	3	0.666667
4562	acetoacetic acid	oil of vitriol	2	3	0.666667
4563	nicotinic acid	oil of vitriol	2	3	0.666667
4564	ascorbic acid	oil of vitriol	2	3	0.666667
4565	acetoacetic acid	oleic acid	2	3	0.666667
4566	nicotinic acid	oleic acid	2	3	0.666667
4567	ascorbic acid	oleic acid	2	3	0.666667
4568	acetoacetic acid	oxyacid	2	3	0.666667
4569	nicotinic acid	oxyacid	2	3	0.666667

4570	ascorbic acid	oxyacid	2	3	0.666667
4571	acetoacetic acid	paba	2	3	0.666667
4572	nicotinic acid	paba	2	3	0.666667
4573	ascorbic acid	paba	2	3	0.666667
4574	acetoacetic acid	phenol	2	3	0.666667
4575	nicotinic acid	phenol	2	3	0.666667
4576	ascorbic acid	phenol	2	3	0.666667
4577	acetoacetic acid	phosphorous acid	2	3	0.666667
4578	nicotinic acid	phosphorous acid	2	3	0.666667
4579	ascorbic acid	phosphorous acid	2	3	0.666667
4580	acetoacetic acid	phthalic acid	2	3	0.666667
4581	nicotinic acid	phthalic acid	2	3	0.666667
4582	ascorbic acid	phthalic acid	2	3	0.666667
4583	acetoacetic acid	racemic acid	2	3	0.666667
4584	nicotinic acid	racemic acid	2	3	0.666667
4585	ascorbic acid	racemic acid	2	3	0.666667
4586	acetoacetic acid	salicylate	2	3	0.666667
4587	nicotinic acid	salicylate	2	3	0.666667
4588	ascorbic acid	salicylate	2	3	0.666667
4589	acetoacetic acid	hydrochloric acid	2	3	0.666667
4590	nicotinic acid	hydrochloric acid	2	3	0.666667
4591	ascorbic acid	hydrochloric acid	2	3	0.666667
4592	acetoacetic acid	ethanedioic acid	2	3	0.666667
4593	nicotinic acid	ethanedioic acid	2	3	0.666667
4594	ascorbic acid	ethanedioic acid	2	3	0.666667
4595	acetoacetic acid	fatty acid	2	3	0.666667
4596	nicotinic acid	fatty acid	2	3	0.666667
4597	ascorbic acid	fatty acid	2	3	0.666667
4598	acetoacetic acid	formic acid	2	3	0.666667
4599	nicotinic acid	formic acid	2	3	0.666667
4600	ascorbic acid	formic acid	2	3	0.666667
4601	acetoacetic acid	gallic acid	2	3	0.666667
4602	nicotinic acid	gallic acid	2	3	0.666667
4603	ascorbic acid	gallic acid	2	3	0.666667
4604	acetoacetic acid	glyceric acid	2	3	0.666667
4605	nicotinic acid	glyceric acid	2	3	0.666667
4606	ascorbic acid	glyceric acid	2	3	0.666667
4607	acetoacetic acid	heptadecanoic acid	2	3	0.666667
4608	nicotinic acid	heptadecanoic acid	2	3	0.666667
4609	ascorbic acid	heptadecanoic acid	2	3	0.666667
4610	acetoacetic acid	cerotic acid	2	3	0.666667
4611	nicotinic acid	cerotic acid	2	3	0.666667
4612	ascorbic acid	cerotic acid	2	3	0.666667
4613	acetoacetic acid	chloroacetic acid	2	3	0.666667
4614	nicotinic acid	chloroacetic acid	2	3	0.666667

4615	ascorbic acid	chloroacetic acid	2	3	0.666667
4616	acetoacetic acid	chromate	2	3	0.666667
4617	nicotinic acid	chromate	2	3	0.666667
4618	ascorbic acid	chromate	2	3	0.666667
4619	acetoacetic acid	chromic acid	2	3	0.666667
4620	nicotinic acid	chromic acid	2	3	0.666667
4621	ascorbic acid	chromic acid	2	3	0.666667
4622	acetoacetic acid	cyanamide	2	3	0.666667
4623	nicotinic acid	cyanamide	2	3	0.666667
4624	ascorbic acid	cyanamide	2	3	0.666667
4625	acetoacetic acid	vinegar	2	3	0.666667
4626	nicotinic acid	vinegar	2	3	0.666667
4627	ascorbic acid	vinegar	2	3	0.666667
4628	acetoacetic acid	acetyl	2	3	0.666667
4629	nicotinic acid	acetyl	2	3	0.666667
4630	ascorbic acid	acetyl	2	3	0.666667
4631	acetoacetic acid	acidophilic	2	3	0.666667
4632	nicotinic acid	acidophilic	2	3	0.666667
4633	ascorbic acid	acidophilic	2	3	0.666667
4634	acetoacetic acid	acidulous	2	3	0.666667
4635	nicotinic acid	acidulous	2	3	0.666667
4636	ascorbic acid	acidulous	2	3	0.666667
4637	acetoacetic acid	arsenate	2	3	0.666667
4638	nicotinic acid	arsenate	2	3	0.666667
4639	ascorbic acid	arsenate	2	3	0.666667
4640	acetoacetic acid	phosphate	2	3	0.666667
4641	nicotinic acid	phosphate	2	3	0.666667
4642	ascorbic acid	phosphate	2	3	0.666667
4643	acetoacetic acid	vitriol	2	3	0.666667
4644	nicotinic acid	vitriol	2	3	0.666667
4645	ascorbic acid	vitriol	2	3	0.666667
4646	acetoacetic acid	selenic acid	2	3	0.666667
4647	nicotinic acid	selenic acid	2	3	0.666667
4648	ascorbic acid	selenic acid	2	3	0.666667
4649	acetoacetic acid	picric acid	2	3	0.666667
4650	nicotinic acid	picric acid	2	3	0.666667
4651	ascorbic acid	picric acid	2	3	0.666667
4652	acetoacetic acid	rna	2	3	0.666667
4653	nicotinic acid	rna	2	3	0.666667
4654	ascorbic acid	rna	2	3	0.666667
4655	acetoacetic acid	aqua fortis	2	3	0.666667
4656	nicotinic acid	aqua fortis	2	3	0.666667
4657	ascorbic acid	aqua fortis	2	3	0.666667
4658	acetoacetic acid	etch	2	3	0.666667
4659	nicotinic acid	etch	2	3	0.666667

4660	ascorbic acid	etch	2	3	0.666667
4661	acetoacetic acid	nitrate	2	3	0.666667
4662	nicotinic acid	nitrate	2	3	0.666667
4663	ascorbic acid	nitrate	2	3	0.666667
4664	acetoacetic acid	superphosphate	2	3	0.666667
4665	nicotinic acid	superphosphate	2	3	0.666667
4666	ascorbic acid	superphosphate	2	3	0.666667
4667	acetoacetic acid	acetify	2	3	0.666667
4668	nicotinic acid	acetify	2	3	0.666667
4669	ascorbic acid	acetify	2	3	0.666667
4670	acetoacetic acid	acidify	2	3	0.666667
4671	nicotinic acid	acidify	2	3	0.666667
4672	ascorbic acid	acidify	2	3	0.666667
4673	acetoacetic acid	acidity	2	3	0.666667
4674	nicotinic acid	acidity	2	3	0.666667
4675	ascorbic acid	acidity	2	3	0.666667
4676	acetoacetic acid	boric acid	2	3	0.666667
4677	nicotinic acid	boric acid	2	3	0.666667
4678	ascorbic acid	boric acid	2	3	0.666667
4679	acetoacetic acid	butyric acid	2	3	0.666667
4680	nicotinic acid	butyric acid	2	3	0.666667
4681	ascorbic acid	butyric acid	2	3	0.666667
4682	acetoacetic acid	citrate	2	3	0.666667
4683	nicotinic acid	citrate	2	3	0.666667
4684	ascorbic acid	citrate	2	3	0.666667
4685	acetoacetic acid	cyanic acid	2	3	0.666667
4686	nicotinic acid	cyanic acid	2	3	0.666667
4687	ascorbic acid	cyanic acid	2	3	0.666667
4688	acetoacetic acid	linolenic acid	2	3	0.666667
4689	nicotinic acid	linolenic acid	2	3	0.666667
4690	ascorbic acid	linolenic acid	2	3	0.666667
4691	acetoacetic acid	fumaric acid	2	3	0.666667
4692	nicotinic acid	fumaric acid	2	3	0.666667
4693	ascorbic acid	fumaric acid	2	3	0.666667
4694	acetoacetic acid	oxalic acid	2	3	0.666667
4695	nicotinic acid	oxalic acid	2	3	0.666667
4696	ascorbic acid	oxalic acid	2	3	0.666667
4697	acetoacetic acid	saccharic acid	2	3	0.666667
4698	nicotinic acid	saccharic acid	2	3	0.666667
4699	ascorbic acid	saccharic acid	2	3	0.666667
4700	acetoacetic acid	sulphate	2	3	0.666667
4701	nicotinic acid	sulphate	2	3	0.666667
4702	ascorbic acid	sulphate	2	3	0.666667
4703	ketone body	senate	3	3	1
4704	legislature	senate	2	3	0.666667

4705	ketone body	house	3	3	1
4706	legislature	house	2	3	0.666667
4707	ketone body	sound	3	3	1
4708	schuller-christian disease	sound	2	3	0.666667
4709	ketone body	propanone	3	3	1
4710	acetone	propanone	2	3	0.666667
4711	ketone body	glycose	2	3	0.666667
4712	glucose tolerance test	glycose	2	3	0.666667
4713	glucose	glycose	1	3	0.333333
4714	ketone body	corpse	3	3	1
4715	recombinant human insulin	corpse	2	3	0.666667
4716	brittle diabetes	short	2	3	0.666667
4717	off	short	2	3	0.666667
4718	light	short	2	3	0.666667
4719	lite	black	2	3	0.666667
4720	light	black	2	3	0.666667
4721	abstemious	black	2	3	0.666667
4722	ketone body	starvation acidosis	3	3	1
4723	diabetic acidosis	starvation acidosis	2	3	0.666667
4724	diabetic diet	dieting	2	3	0.666667
4725	light diet	dieting	2	3	0.666667
4726	macrobiotic diet	dieting	2	3	0.666667
4727	diabetic diet	legislature	2	3	0.666667
4728	light diet	legislature	2	3	0.666667
4729	macrobiotic diet	legislature	2	3	0.666667
4730	cushing's disease	pip	2	3	0.666667
4731	hand-schuller-christian disease	pip	2	3	0.666667
4732	schuller-christian disease	pip	2	3	0.666667
4733	cushing's disease	plague	2	3	0.666667
4734	hand-schuller-christian disease	plague	2	3	0.666667
4735	schuller-christian disease	plague	2	3	0.666667

4736	cushing's disease	lupus	2	3	0.666667
4737	hand-schuller-christian disease	lupus	2	3	0.666667
4738	schuller-christian disease	lupus	2	3	0.666667
4739	cushing's disease	influenza	2	3	0.666667
4740	hand-schuller-christian disease	influenza	2	3	0.666667
4741	schuller-christian disease	influenza	2	3	0.666667
4742	hand-schuller-christian disease	get	2	3	0.666667
4743	insulin shock treatment	get	1	3	0.333333
4744	off	get	2	3	0.666667
4745	fare	get	1	3	0.333333
4746	acetoacetic acid	aspirin	2	3	0.666667
4747	nicotinic acid	aspirin	2	3	0.666667
4748	ascorbic acid	aspirin	2	3	0.666667
4749	sir frederick grant banting	brown	2	3	0.666667
4750	john james rickard macleod	brown	2	3	0.666667
4751	john macleod	brown	2	3	0.666667
4752	antidiabetic drug	abortifacient	2	3	0.666667
4753	hypoglycaemic agent	abortifacient	2	3	0.666667
4754	hypoglycemic agent	abortifacient	2	3	0.666667
4755	antidiabetic drug	depressant	2	3	0.666667
4756	hypoglycaemic agent	depressant	2	3	0.666667
4757	hypoglycemic agent	depressant	2	3	0.666667
4758	antidiabetic drug	anodyne	2	3	0.666667
4759	reducing diet	anodyne	2	3	0.666667
4760	botanical medicine	anodyne	2	3	0.666667
4761	acetoacetic acid	hydrogen cyanide	2	3	0.666667

4762	nicotinic acid	hydrogen cyanide	2	3	0.666667
4763	ascorbic acid	hydrogen cyanide	2	3	0.666667
4764	sir frederick grant banting	ives	2	3	0.666667
4765	frederick sanger	ives	2	3	0.666667
4766	charles herbert best	ives	2	3	0.666667
4767	sir frederick grant banting	evans	2	3	0.666667
4768	john james rickard macleod	evans	2	3	0.666667
4769	charles herbert best	evans	2	3	0.666667
4770	sir frederick grant banting	parry	2	3	0.666667
4771	off	parry	2	3	0.666667
4772	charles herbert best	parry	1	3	0.333333
4773	avoid	parry	1	3	0.333333
4774	sir frederick grant banting	churchill	2	3	0.666667
4775	john james rickard macleod	churchill	2	3	0.666667
4776	john macleod	churchill	2	3	0.666667
4777	mature-onset diabetes	make	1	3	0.333333
4778	best	make	1	3	0.333333
4779	spare	make	1	3	0.333333
4780	adult-onset diabetes	character	2	3	0.666667
4781	type i diabetes	character	2	3	0.666667
4782	adult-onset diabetes mellitus	character	2	3	0.666667
4783	adult-onset diabetes	catch	2	3	0.666667
4784	adult-onset diabetes mellitus	catch	2	3	0.666667
4785	hand-schuller-christian disease	catch	2	3	0.666667
4786	adult-onset diabetes	counterattack	2	3	0.666667
4787	growth-onset diabetes	counterattack	2	3	0.666667
4788	mature-onset diabetes	counterattack	2	3	0.666667

4789	adult-onset diabetes	diversion	2	3	0.666667
4790	growth-onset diabetes	diversion	2	3	0.666667
4791	mature-onset diabetes	diversion	2	3	0.666667
4792	adult-onset diabetes	incursion	2	3	0.666667
4793	growth-onset diabetes	incursion	2	3	0.666667
4794	mature-onset diabetes	incursion	2	3	0.666667
4795	adult-onset diabetes	banzai attack	2	3	0.666667
4796	growth-onset diabetes	banzai attack	2	3	0.666667
4797	mature-onset diabetes	banzai attack	2	3	0.666667
4798	adult-onset diabetes	bombardment	2	3	0.666667
4799	growth-onset diabetes	bombardment	2	3	0.666667
4800	mature-onset diabetes	bombardment	2	3	0.666667
4801	adult-onset diabetes	penetration	2	3	0.666667
4802	growth-onset diabetes	penetration	2	3	0.666667
4803	mature-onset diabetes	penetration	2	3	0.666667
4804	adult-onset diabetes	invasion	2	3	0.666667
4805	growth-onset diabetes	invasion	2	3	0.666667
4806	mature-onset diabetes	invasion	2	3	0.666667
4807	adult-onset diabetes	girl	2	3	0.666667
4808	adult-onset diabetes mellitus	girl	2	3	0.666667
4809	juvenile	girl	2	3	0.666667
4810	ketosis-resistant diabetes	ketotic	2	3	0.666667
4811	ketosis-prone diabetes	ketotic	2	3	0.666667
4812	ketosis-resistant	ketotic	2	3	0.666667

	diabetes mellitus				
4813	growth-onset diabetes	check	2	3	0.666667
4814	train	check	2	3	0.666667
4815	off	check	2	3	0.666667
4816	growth-onset diabetes	prepuberty	2	3	0.666667
4817	maturity-onset diabetes	prepuberty	2	3	0.666667
4818	maturity-onset diabetes mellitus	prepuberty	2	3	0.666667
4819	type ii diabetes	run	1	3	0.333333
4820	off	run	3	3	1
4821	ulcer diet	run	1	3	0.333333
4822	follow	run	1	3	0.333333
4823	type ii diabetes	classic	2	3	0.666667
4824	best	classic	1	3	0.333333
4825	charles herbert best	classic	1	3	0.333333
4826	type ii diabetes	flower	2	3	0.666667
4827	best	flower	2	3	0.666667
4828	non-insulin-dependent diabetes	mass	2	3	0.666667
4829	ketone body	mass	3	3	1
4830	maturity-onset diabetes	adult	2	3	0.666667
4831	maturity-onset diabetes mellitus	adult	2	3	0.666667
4832	fare	adult	2	3	0.666667
4833	non-insulin-dependent diabetes	absolute	2	3	0.666667
4834	non-insulin-dependent diabetes mellitus	absolute	2	3	0.666667
4835	arbitrary	absolute	2	3	0.666667
4836	non-insulin-dependent diabetes	condition	2	2	1
4837	non-insulin-dependent diabetes mellitus	condition	1	2	0.5
4838	train	condition	1	2	0.5

4839	clear liquid diet	jejune	1	2	0.5
4840	liquid diet	jejune	2	2	1
4841	bland diet	jejune	1	2	0.5
4842	maturity-onset diabetes	prime	1	2	0.5
4843	maturity-onset diabetes mellitus	prime	1	2	0.5
4844	best	prime	1	2	0.5
4845	maturity-onset diabetes	grow	2	2	1
4846	maturity-onset diabetes mellitus	grow	1	2	0.5
4847	nurture	grow	1	2	0.5
4848	sir frederick grant banting	page	2	2	1
4849	frederick sanger	page	1	2	0.5
4850	legislature	page	1	2	0.5
4851	type ii diabetes	block	2	2	1
4852	iron overload	block	1	2	0.5
4853	off	block	1	2	0.5
4854	maturity-onset diabetes	come	1	2	0.5
4855	maturity-onset diabetes mellitus	come	1	2	0.5
4856	follow	come	1	2	0.5
4857	fare	come	1	2	0.5
4858	adult-onset diabetes	man	2	2	1
4859	adult-onset diabetes mellitus	man	1	2	0.5
4860	isles of langerhans	man	1	2	0.5
4861	type i diabetes	drag	2	2	1
4862	type ii diabetes	drag	1	2	0.5
4863	iron overload	drag	1	2	0.5
4864	type i diabetes	order	1	2	0.5
4865	hand-schuller-christian disease	order	2	2	1
4866	schuller-christian disease	order	1	2	0.5
4867	antidiabetic drug	anticonvulsant	2	2	1

4868	diabetic coma	anticonvulsant	1	2	0.5
4869	botanical medicine	anticonvulsant	1	2	0.5
4870	carbohydrate loading	complex	1	2	0.5
4871	hand-schuller-christian disease	complex	1	2	0.5
4872	schuller-christian disease	complex	1	2	0.5
4873	carbohydrate	complex	1	2	0.5
4874	cushing's disease	communicate	1	2	0.5
4875	hand-schuller-christian disease	communicate	1	2	0.5
4876	schuller-christian disease	communicate	2	2	1
4877	diabetic diet	antiphlogistic	1	2	0.5
4878	reducing diet	antiphlogistic	1	2	0.5
4879	macrobiotic diet	antiphlogistic	2	2	1
4880	diabetic diet	beneficial	1	2	0.5
4881	balanced diet	beneficial	1	2	0.5
4882	macrobiotic diet	beneficial	2	2	1
4883	diabetic diet	arbitrary	1	2	0.5
4884	low-calorie diet	arbitrary	1	2	0.5
4885	macrobiotic diet	arbitrary	2	2	1
4886	diabetic diet	carbohydrate	1	1	1
4887	macrobiotic diet	carbohydrate	1	1	1
4888	ketone body	right	1	2	0.5
4889	hand-schuller-christian disease	right	1	2	0.5
4890	insulin shock treatment	right	1	2	0.5
4891	diabetic coma	planet	1	2	0.5
4892	ketone body	planet	1	2	0.5
4893	kussmaul's coma	planet	1	2	0.5
4894	ketone body	star	1	2	0.5
4895	lite	star	1	2	0.5
4896	abstemious	star	1	2	0.5

4897	iron-storage disease	rust	1	2	0.5
4898	schuller-christian disease	rust	1	2	0.5
4899	eat	rust	2	2	1
4900	iron-storage disease	chlorosis	2	2	1
4901	iron overload	chlorosis	1	2	0.5
4902	vitamin-deficiency diet	chlorosis	1	2	0.5
4903	hypoglycaemic agent	ferment	1	2	0.5
4904	hypoglycemic agent	ferment	1	2	0.5
4905	carbohydrate loading	ferment	1	2	0.5
4906	carbohydrate	ferment	1	2	0.5
4907	beta cell	propranolol	1	2	0.5
4908	hand-schuller-christian disease	propranolol	1	2	0.5
4909	schuller-christian disease	propranolol	2	2	1
4910	glucose tolerance test	cross	1	2	0.5
4911	hand-schuller-christian disease	cross	2	2	1
4912	schuller-christian disease	cross	1	2	0.5
4913	hand-schuller-christian disease	christian science	1	2	0.5
4914	schuller-christian disease	christian science	2	2	1
4915	protestant	christian science	1	2	0.5
4916	fred sanger	back	1	2	0.5
4917	off	back	1	2	0.5
4918	nurture	back	2	2	1
4919	clear liquid diet	drain	1	2	0.5
4920	off	drain	2	2	1
4921	eat	drain	1	2	0.5
4922	clear liquid diet	fast	2	2	1
4923	liquid diet	fast	1	2	0.5
4924	eat	fast	1	2	0.5

4925	soft	easy	1	2	0.5
4926	soft diet	easy	2	2	1
4927	light	easy	1	2	0.5
4928	low-salt diet	sodium	2	2	1
4929	salt-free diet	sodium	1	2	0.5
4930	soft diet	sodium	1	2	0.5
4931	spoon food	dip	1	2	0.5
4932	eat	dip	1	2	0.5
4933	fare	dip	2	2	1
4934	bland	table	1	2	0.5
4935	spoon food	table	2	2	1
4936	fare	table	1	2	0.5
4937	fad diet	course	2	2	1
4938	regimen	course	1	2	0.5
4939	follow	course	1	2	0.5
4940	low-calorie diet	weak	2	2	1
4941	low-fat diet	weak	1	2	0.5
4942	light	weak	1	2	0.5
4943	carbohydrate loading	cane sugar	1	2	0.5
4944	glucose	cane sugar	1	2	0.5
4945	carbohydrate	cane sugar	2	2	1
4946	high-vitamin diet	deficiency+disease	1	2	0.5
4947	vitamin-deficiency diet	deficiency+disease	2	2	1
4948	dietary	deficiency+disease	1	2	0.5
4949	carbohydrate loading	sucrate	2	2	1
4950	low-sodium diet	sucrate	1	2	0.5
4951	carbohydrate	sucrate	1	2	0.5
4952	high-protein diet	porter	1	2	0.5
4953	high-vitamin diet	porter	1	2	0.5
4954	charles herbert best	porter	1	2	0.5
4955	best	dog	1	2	0.5
4956	follow	dog	1	2	0.5
4957	best	place	1	2	0.5
4958	train	place	1	2	0.5
4959	well-fed	feed	1	2	0.5
4960	eat	feed	1	2	0.5
4961	john james rickard macleod	napier	2	2	1
4962	charles herbert best	napier	1	2	0.5

4963	john macleod	napier	1	2	0.5
4964	insulin reaction	reducing	1	2	0.5
4965	reduce	reducing	2	2	1
4966	dieting	reducing	1	2	0.5
4967	high-vitamin diet	pantothenic acid	1	2	0.5
4968	vitamin-deficiency diet	pantothenic acid	1	2	0.5
4969	amino acid	pantothenic acid	1	2	0.5
4970	fred sanger	tail	1	2	0.5
4971	nurture	tail	2	2	1
4972	follow	tail	1	2	0.5
4973	high-protein diet	gluten	1	2	0.5
4974	clear liquid diet	gluten	2	2	1
4975	salt-free diet	gluten	1	2	0.5
4976	clear liquid diet	geophagy	1	2	0.5
4977	liquid diet	geophagy	2	2	1
4978	vitamin-deficiency diet	geophagy	1	2	0.5
4979	lite	shadow	2	2	1
4980	abstemious	shadow	1	2	0.5
4981	follow	shadow	1	2	0.5
4982	insulin shock	horrify	2	2	1
4983	insulin shock treatment	horrify	1	2	0.5
4984	lite	white	2	2	1
4985	light	white	1	2	0.5
4986	abstemious	white	1	2	0.5
4987	lite	aurora	1	1	1
4988	abstemious	aurora	1	1	1
4989	lite	blind	1	1	1
4990	abstemious	blind	1	1	1
4991	lite	dazzle	1	1	1
4992	abstemious	dazzle	1	1	1
4993	lite	exposure	1	1	1
4994	abstemious	exposure	1	1	1
4995	lite	floodlight	1	1	1
4996	abstemious	floodlight	1	1	1
4997	lite	fluorescence	1	1	1
4998	abstemious	fluorescence	1	1	1
4999	lite	gig	1	1	1
5000	abstemious	gig	1	1	1
5001	lite	glare	1	1	1
5002	abstemious	glare	1	1	1
5003	lite	glimmer	1	1	1

5004	abstemious	glimmer	1	1	1
5005	lite	glitter	1	1	1
5006	abstemious	glitter	1	1	1
5007	lite	kite	1	1	1
5008	abstemious	kite	1	1	1
5009	lite	lantern	1	1	1
5010	abstemious	lantern	1	1	1
5011	lite	luminous	1	1	1
5012	abstemious	luminous	1	1	1
5013	lite	moon	1	1	1
5014	abstemious	moon	1	1	1
5015	lite	moonlight	1	1	1
5016	abstemious	moonlight	1	1	1
5017	lite	opaque	1	1	1
5018	abstemious	opaque	1	1	1
5019	lite	photic	1	1	1
5020	abstemious	photic	1	1	1
5021	lite	radiant	1	1	1
5022	abstemious	radiant	1	1	1
5023	lite	reflect	1	1	1
5024	abstemious	reflect	1	1	1
5025	lite	sidelight	1	1	1
5026	abstemious	sidelight	1	1	1
5027	lite	spotlight	1	1	1
5028	abstemious	spotlight	1	1	1
5029	lite	twinkle	1	1	1
5030	abstemious	twinkle	1	1	1
5031	lite	wherry	1	1	1
5032	abstemious	wherry	1	1	1
5033	islets of langerhans	cays	1	1	1
5034	isles of langerhans	cays	1	1	1
5035	islets of langerhans	continental	1	1	1
5036	isles of langerhans	continental	1	1	1
5037	follow	go	1	1	1
5038	fare	go	1	1	1
5039	follow	clerk	1	1	1
5040	legislature	clerk	1	1	1
5041	well-fed	carry	1	1	1
5042	nurture	carry	1	1	1
5043	supplement	accompany	1	1	1
5044	follow	accompany	1	1	1

5045	train	coach	1	1	1
5046	athletic training	coach	1	1	1
5047	train	groom	1	1	1
5048	well-fed	groom	1	1	1
5049	train	educate	1	1	1
5050	nurture	educate	1	1	1
5051	train	cultivate	1	1	1
5052	nurture	cultivate	1	1	1
5053	train	discipline	1	1	1
5054	athletic training	discipline	1	1	1
5055	train	drill	1	1	1
5056	athletic training	drill	1	1	1
5057	train	exercise	1	1	1
5058	athletic training	exercise	1	1	1
5059	clear liquid diet	bubble	1	1	1
5060	liquid diet	bubble	1	1	1
5061	clear liquid diet	rinse	1	1	1
5062	liquid diet	rinse	1	1	1
5063	low-fat diet	chyle	1	1	1
5064	liquid diet	chyle	1	1	1
5065	clear liquid diet	ether	1	1	1
5066	liquid diet	ether	1	1	1
5067	low-fat diet	salt pork	1	1	1
5068	low-salt diet	salt pork	1	1	1
5069	carbo loading	bolt	1	1	1
5070	eat	bolt	1	1	1
5071	carbo loading	freight	1	1	1
5072	train	freight	1	1	1
5073	carbo loading	freight liner	1	1	1
5074	train	freight liner	1	1	1
5075	spare	save	1	1	1
5076	avoid	save	1	1	1
5077	carbo loading	conducive	1	1	1
5078	carbohydrate loading	conducive	1	1	1
5079	follow	result	1	1	1
5080	beneficial	result	1	1	1
5081	high-protein diet	laser	1	1	1
5082	high-vitamin diet	laser	1	1	1
5083	clear liquid diet	dry	1	1	1
5084	salt-free diet	dry	1	1	1
5085	clear liquid diet	vary	1	1	1
5086	light diet	vary	1	1	1

5087	train	guard	1	1	1
5088	follow	guard	1	1	1
5089	bland	board	1	1	1
5090	fare	board	1	1	1
5091	bland	blandly	1	1	1
5092	bland diet	blandly	1	1	1
5093	bland	blandness	1	1	1
5094	bland diet	blandness	1	1	1
5095	bland	insipid	1	1	1
5096	bland diet	insipid	1	1	1
5097	bland	suave	1	1	1
5098	bland diet	suave	1	1	1
5099	bland	blanding	1	1	1
5100	bland diet	blanding	1	1	1
5101	bland	flavorless	1	1	1
5102	bland diet	flavorless	1	1	1
5103	bland	flavourless	1	1	1
5104	bland diet	flavourless	1	1	1
5105	bland	savorless	1	1	1
5106	bland diet	savorless	1	1	1
5107	bland	vapid	1	1	1
5108	bland diet	vapid	1	1	1
5109	bland	vanilla	1	1	1
5110	bland diet	vanilla	1	1	1
5111	bland	arbutus unedo	1	1	1
5112	bland diet	arbutus unedo	1	1	1
5113	bland	bromide	1	1	1
5114	bland diet	bromide	1	1	1
5115	bland	cottage+cheese	1	1	1
5116	bland diet	cottage+cheese	1	1	1
5117	bland	drastic	1	1	1
5118	bland diet	drastic	1	1	1
5119	bland	faair-spoken	1	1	1
5120	bland diet	faair-spoken	1	1	1
5121	bland	favonian	1	1	1
5122	bland diet	favonian	1	1	1
5123	bland	impregnably	1	1	1
5124	bland diet	impregnably	1	1	1
5125	bland	irish strawberry	1	1	1
5126	bland diet	irish strawberry	1	1	1
5127	bland	muenster	1	1	1
5128	bland diet	muenster	1	1	1
5129	bland	obtundent	1	1	1
5130	bland diet	obtundent	1	1	1
5131	bland	pop	1	1	1

5132	bland diet	pop	1	1	1
5133	bland	pop music	1	1	1
5134	bland diet	pop music	1	1	1
5135	bland	strawberry tree	1	1	1
5136	bland diet	strawberry tree	1	1	1
5137	bland	tame	1	1	1
5138	bland diet	tame	1	1	1
5139	bland	tasteless	1	1	1
5140	bland diet	tasteless	1	1	1
5141	bland	til seed	1	1	1
5142	bland diet	til seed	1	1	1
5143	bland	unctuous	1	1	1
5144	bland diet	unctuous	1	1	1
5145	bland	watery	1	1	1
5146	bland diet	watery	1	1	1
5147	legislature	conference	1	1	1
5148	athletic training	conference	1	1	1
5149	recombinant human insulin	gland	1	1	1
5150	secretes	gland	1	1	1
5151	fred sanger	rearward	1	1	1
5152	nurture	rearward	1	1	1
5153	islands of langerhans	hebrides	1	1	1
5154	isles of langerhans	hebrides	1	1	1
5155	high-protein diet	electron	1	1	1
5156	high-vitamin diet	electron	1	1	1
5157	insulin shock	distributive shock	1	1	1
5158	insulin shock	obstructive shock	1	1	1
5159	insulin shock	slam	1	1	1
5160	insulin shock	fright	1	1	1
5161	insulin shock	shake up	1	1	1
5162	insulin shock	cardiogenic shock	1	1	1
5163	insulin shock	earthquake	1	1	1
5164	insulin shock	electric	1	1	1
5165	insulin shock	hormephobia	1	1	1
5166	insulin shock	hypovolemic shock	1	1	1
5167	insulin shock therapy	torpillage	1	1	1
5168	insulin shock treatment	torpillage	1	1	1
5169	insulin shock therapy	metrazol	1	1	1

5170	insulin shock treatment	metrazol	1	1	1
5171	insulin shock therapy	pentamethylenetetrazol	1	1	1
5172	insulin shock treatment	pentamethylenetetrazol	1	1	1
5173	insulin shock therapy	pentylenetetrazol	1	1	1
5174	insulin shock treatment	pentylenetetrazol	1	1	1
5175	insulin shock therapy	sensitive	1	1	1
5176	insulin shock treatment	sensitive	1	1	1
5177	insulin shock therapy	defibrillation	1	1	1
5178	insulin shock treatment	defibrillation	1	1	1
5179	insulin shock therapy	serum albumin	1	1	1
5180	insulin shock treatment	serum albumin	1	1	1
5181	insulin shock therapy	electroshock+therapy	1	1	1
5182	insulin shock treatment	electroshock+therapy	1	1	1
5183	insulin shock therapy	galvanism	1	1	1
5184	insulin shock treatment	galvanism	1	1	1
5185	insulin shock therapy	shock treatment	1	1	1
5186	insulin shock treatment	shock treatment	1	1	1
5187	high-protein diet	merry	1	1	1
5188	high-vitamin diet	merry	1	1	1
5189	high-protein diet	cruiser	1	1	1
5190	high-vitamin diet	cruiser	1	1	1
5191	high-protein diet	badminton	1	1	1
5192	high-vitamin diet	badminton	1	1	1
5193	high-protein diet	cot	1	1	1
5194	high-vitamin diet	cot	1	1	1

5195	high-protein diet	strobe lighting	1	1	1
5196	high-vitamin diet	strobe lighting	1	1	1
5197	high-protein diet	top+boot	1	1	1
5198	high-vitamin diet	top+boot	1	1	1
5199	high-protein diet	skyrocket	1	1	1
5200	high-vitamin diet	skyrocket	1	1	1
5201	high-protein diet	hydroplane	1	1	1
5202	high-vitamin diet	hydroplane	1	1	1
5203	high-protein diet	spallation	1	1	1
5204	high-vitamin diet	spallation	1	1	1
5205	high-protein diet	horse latitudes	1	1	1
5206	high-vitamin diet	horse latitudes	1	1	1
5207	high-protein diet	synchrotron radiation	1	1	1
5208	high-vitamin diet	synchrotron radiation	1	1	1
5209	high-protein diet	polymer	1	1	1
5210	high-vitamin diet	polymer	1	1	1
5211	high-protein diet	incandescence	1	1	1
5212	high-vitamin diet	incandescence	1	1	1
5213	high-protein diet	incandescent	1	1	1
5214	high-vitamin diet	incandescent	1	1	1
5215	high-protein diet	lighthouse	1	1	1
5216	high-vitamin diet	lighthouse	1	1	1
5217	high-protein diet	top-boots	1	1	1
5218	high-vitamin diet	top-boots	1	1	1
5219	high-protein diet	white-hot	1	1	1

5220	high-vitamin diet	white-hot	1	1	1
5221	high-protein diet	mao jacket	1	1	1
5222	high-vitamin diet	mao jacket	1	1	1
5223	high-protein diet	spider phaeton	1	1	1
5224	high-vitamin diet	spider phaeton	1	1	1
5225	high-protein diet	horse+latitudes	1	1	1
5226	high-vitamin diet	horse+latitudes	1	1	1
5227	high-protein diet	process plate	1	1	1
5228	high-vitamin diet	process plate	1	1	1
5229	high-protein diet	relax	1	1	1
5230	high-vitamin diet	relax	1	1	1
5231	high-protein diet	fusion bomb	1	1	1
5232	high-vitamin diet	fusion bomb	1	1	1
5233	high-protein diet	h-bomb	1	1	1
5234	high-vitamin diet	h-bomb	1	1	1
5235	high-protein diet	hydrogen bomb	1	1	1
5236	high-vitamin diet	hydrogen bomb	1	1	1
5237	high-protein diet	thermonuclear bomb	1	1	1
5238	high-vitamin diet	thermonuclear bomb	1	1	1
5239	high-protein diet	phase-contrast microscope	1	1	1
5240	high-vitamin diet	phase-contrast microscope	1	1	1
5241	high-protein diet	krypton	1	1	1
5242	high-vitamin diet	krypton	1	1	1
5243	high-protein diet	express rifle	1	1	1
5244	high-vitamin diet	express rifle	1	1	1

5245	high-protein diet	aluminum	1	1	1
5246	high-vitamin diet	aluminum	1	1	1
5247	high-protein diet	joy	1	1	1
5248	high-vitamin diet	joy	1	1	1
5249	high-protein diet	kerr cell	1	1	1
5250	high-vitamin diet	kerr cell	1	1	1
5251	high-protein diet	els	1	1	1
5252	high-vitamin diet	els	1	1	1
5253	high-protein diet	zip	1	1	1
5254	high-vitamin diet	zip	1	1	1
5255	amino acid	codon	1	1	1
5256	amino acid	dopa	1	1	1
5257	amino acid	isoleucine	1	1	1
5258	amino acid	threonine	1	1	1
5259	amino acid	glutamine	1	1	1
5260	amino acid	essential amino acid	1	1	1
5261	amino acid	lysine	1	1	1
5262	amino acid	gaba	1	1	1
5263	amino acid	gamma aminobutyric acid	1	1	1
5264	amino acid	glutaminic acid	1	1	1
5265	amino acid	iodoamino acid	1	1	1
5266	amino acid	sarcosine	1	1	1
5267	amino acid	transfer rna	1	1	1
5268	amino acid	taurine	1	1	1
5269	amino acid	inessential amino acid	1	1	1
5270	amino acid	degenerate	1	1	1
5271	amino acid	trna	1	1	1
5272	amino acid	genetic code	1	1	1
5273	amino acid	glutathione	1	1	1
5274	amino acid	dihydroxyphenylalanine	1	1	1
5275	amino acid	triiodothyronine	1	1	1
5276	amino acid	peptide bond	1	1	1
5277	amino acid	creatine	1	1	1
5278	amino acid	creatin	1	1	1
5279	amino acid	degeneracy	1	1	1
5280	amino acid	ethionine	1	1	1

5281	amino acid	glutamic+acid	1	1	1
5282	amino acid	pentapeptide	1	1	1
5283	amino acid	isoelectric point	1	1	1
5284	amino acid	tyramine	1	1	1
5285	amino acid	structural+gene	1	1	1
5286	amino acid	translate	1	1	1
5287	amino acid	thyroxine	1	1	1
5288	amino acid	anticodon	1	1	1
5289	amino acid	lysine intolerance	1	1	1
5290	amino acid	nonsense	1	1	1
5291	amino acid	lysinemia	1	1	1
5292	amino acid	triplet code	1	1	1
5293	amino acid	acceptor rna	1	1	1
5294	amino acid	soluble rna	1	1	1
5295	amino acid	ribosome	1	1	1
5296	amino acid	polypeptide	1	1	1
5297	amino acid	aminoalkanoic acid	1	1	1
5298	amino acid	messenger rna	1	1	1
5299	amino acid	diamine	1	1	1
5300	amino acid	peptone	1	1	1
5301	amino acid	sequence	1	1	1
5302	amino acid	deamination	1	1	1
5303	amino acid	deaminization	1	1	1
5304	amino acid	aminopherase	1	1	1
5305	amino acid	aminotransferase	1	1	1
5306	amino acid	carbamino	1	1	1
5307	amino acid	transaminase	1	1	1
5308	amino acid	glutamate	1	1	1
5309	amino acid	linoleic acid	1	1	1
5310	amino acid	aminoplast	1	1	1
5311	amino acid	deaminate	1	1	1
5312	amino acid	linolic acid	1	1	1
5313	amino acid	transaminate	1	1	1
5314	amino acid	transamination	1	1	1
5315	vitamin-deficiency diet	supply	1	1	1
5316	supplement	supply	1	1	1
5317	amino acid	code	1	1	1
5318	vitamin-deficiency diet	hypovitaminosis	1	1	1
5319	avitaminosis	hypovitaminosis	1	1	1
5320	vitamin-deficiency diet	rachitis	1	1	1
5321	avitaminosis	rachitis	1	1	1

5322	high-vitamin diet	tocopherol	1	1	1
5323	vitamin-deficiency diet	tocopherol	1	1	1
5324	high-vitamin diet	night blindness	1	1	1
5325	vitamin-deficiency diet	night blindness	1	1	1
5326	high-vitamin diet	moon blindness	1	1	1
5327	vitamin-deficiency diet	moon blindness	1	1	1
5328	high-vitamin diet	hypothrombinemia	1	1	1
5329	vitamin-deficiency diet	hypothrombinemia	1	1	1
5330	high-vitamin diet	biotin	1	1	1
5331	vitamin-deficiency diet	biotin	1	1	1
5332	high-vitamin diet	cobalamin	1	1	1
5333	vitamin-deficiency diet	cobalamin	1	1	1
5334	high-vitamin diet	cyanocobalamin	1	1	1
5335	vitamin-deficiency diet	cyanocobalamin	1	1	1
5336	high-vitamin diet	antipernicious anemia factor	1	1	1
5337	vitamin-deficiency diet	antipernicious anemia factor	1	1	1
5338	nicotinic acid	niacin	1	1	1
5339	vitamin-deficiency diet	niacin	1	1	1
5340	pellagra	nicotinic+acid	1	1	1
5341	nicotinic acid	nicotinic+acid	1	1	1
5342	salt-free diet	clean	1	1	1
5343	light	clean	1	1	1
5344	salt-free diet	clear	1	1	1
5345	light	clear	1	1	1
5346	high-protein diet	happy	1	1	1
5347	high-vitamin diet	happy	1	1	1
5348	high-protein diet	manograph	1	1	1
5349	high-vitamin diet	manograph	1	1	1

5350	high-protein diet	antibody	1	1	1
5351	high-vitamin diet	antibody	1	1	1
5352	high-protein diet	genus chlorella	1	1	1
5353	high-vitamin diet	genus chlorella	1	1	1
5354	high-protein diet	rocket	1	1	1
5355	high-vitamin diet	rocket	1	1	1
5356	low-salt diet	blue	1	1	1
5357	low-sodium diet	blue	1	1	1
5358	low-salt diet	determination	1	1	1
5359	salt-free diet	determination	1	1	1
5360	low-salt diet	silver	1	1	1
5361	salt-free diet	silver	1	1	1
5362	salt-free diet	simplicity	1	1	1
5363	light diet	simplicity	1	1	1
5364	light diet	recess	1	1	1
5365	ulcer diet	recess	1	1	1
5366	fiber	fibre	1	1	1
5367	roughage	fibre	1	1	1
5368	reduce	reduction	1	1	1
5369	reducing diet	reduction	1	1	1
5370	john macleod	wesleyan	1	1	1
5371	protestant	wesleyan	1	1	1
5372	train	car	1	1	1
5373	charles herbert best	car	1	1	1
5374	john james rickard macleod	convention	1	1	1
5375	arbitrary	convention	1	1	1
5376	john james rickard macleod	wilkes	1	1	1
5377	charles herbert best	wilkes	1	1	1
5378	charles herbert best	wesley	1	1	1
5379	john macleod	wesley	1	1	1
5380	john james rickard macleod	watson	1	1	1
5381	john macleod	watson	1	1	1
5382	john james rickard macleod	marshall	1	1	1
5383	john macleod	marshall	1	1	1

5384	john james rickard macleod	byng	1	1	1
5385	john macleod	byng	1	1	1
5386	spoon food	spot	1	1	1
5387	spoon food	salt	1	1	1
5388	off	shoot	1	1	1
5389	off	mark	1	1	1
5390	spare	mark	1	1	1
5391	insulin shock treatment	work	1	1	1
5392	follow	work	1	1	1
5393	nurture	father	1	1	1
5394	amino acid	sulfur	1	1	1
5395	john james rickard macleod	whig	1	1	1
5396	charles herbert best	whig	1	1	1
5397	john james rickard macleod	eliot	1	1	1
5398	charles herbert best	eliot	1	1	1
5399	john james rickard macleod	stuart	1	1	1
5400	charles herbert best	stuart	1	1	1
5401	john james rickard macleod	canning	1	1	1
5402	charles herbert best	canning	1	1	1
5403	best	hit	1	1	1
5404	best	record	1	1	1
5405	best	tiptop	1	1	1
5406	best	paranymph	1	1	1
5407	best	elite	1	1	1
5408	best	pick	1	1	1
5409	best	second-best	1	1	1
5410	best	chart	1	1	1
5411	charles herbert best	rollo	1	1	1
5412	charles herbert best	snow	1	1	1
5413	best	primary	1	1	1
5414	best	scoop	1	1	1
5415	spoon food	scoop	1	1	1
5416	best	level	1	1	1
5417	train	level	1	1	1
5418	acetoacetic acid	aqua regia	1	1	1

5419	ascorbic acid	aqua regia	1	1	1
5420	cushing's disease	osteopetrosis	1	1	1
5421	schuller-christian disease	osteopetrosis	1	1	1
5422	cushing's disease	herpes	1	1	1
5423	schuller-christian disease	herpes	1	1	1
5424	cushing's disease	albers-schonberg disease	1	1	1
5425	schuller-christian disease	albers-schonberg disease	1	1	1
5426	cushing's disease	hyperadrenalism	1	1	1
5427	obesity diet	hyperadrenalism	1	1	1
5428	best	better	1	1	1
5429	best	cream	1	1	1
5430	best	optimum	1	1	1
5431	best	aristocracy	1	1	1
5432	carbohydrate loading	triticin	1	1	1
5433	carbohydrate	triticin	1	1	1
5434	glucose	candy	1	1	1
5435	carbohydrate	candy	1	1	1
5436	glucose	praline	1	1	1
5437	carbohydrate	praline	1	1	1
5438	carbohydrate loading	saccharate	1	1	1
5439	carbohydrate	saccharate	1	1	1
5440	carbohydrate loading	saccharification	1	1	1
5441	carbohydrate	saccharification	1	1	1
5442	carbohydrate loading	saccharify	1	1	1
5443	carbohydrate	saccharify	1	1	1
5444	carbohydrate loading	saccharose	1	1	1
5445	carbohydrate	saccharose	1	1	1
5446	carbohydrate loading	sinistrin	1	1	1
5447	carbohydrate	sinistrin	1	1	1
5448	dietary	kelp	1	1	1
5449	supplement	kelp	1	1	1
5450	dietary	nutraceutical	1	1	1

5451	supplement	nutraceutical	1	1	1
5452	dietary	vitamin pill	1	1	1
5453	supplement	vitamin pill	1	1	1
5454	recombinant human insulin	liver	1	1	1
5455	secretes	liver	1	1	1
5456	dietary	mannitol	1	1	1
5457	supplement	mannitol	1	1	1
5458	vitamin-deficiency diet	creeps	1	1	1
5459	dietary	creeps	1	1	1
5460	amino acid	nonessential	1	1	1
5461	soft	mash	1	1	1
5462	pap	mash	1	1	1
5463	bland	flummery	1	1	1
5464	bland diet	flummery	1	1	1
5465	low-fat diet	adipocere	1	1	1
5466	soft diet	adipocere	1	1	1
5467	soft	fluff	1	1	1
5468	soft diet	fluff	1	1	1
5469	soft	fluffy	1	1	1
5470	soft diet	fluffy	1	1	1
5471	soft	poplar	1	1	1
5472	soft diet	poplar	1	1	1
5473	soft	flannel	1	1	1
5474	soft diet	flannel	1	1	1
5475	carbohydrate loading	levulosan	1	1	1
5476	carbohydrate	levulosan	1	1	1
5477	carbohydrate loading	maple sugar	1	1	1
5478	carbohydrate	maple sugar	1	1	1
5479	carbohydrate loading	mucilage	1	1	1
5480	carbohydrate	mucilage	1	1	1
5481	carbohydrate loading	ribose	1	1	1
5482	carbohydrate	ribose	1	1	1
5483	carbohydrate loading	saccharide	1	1	1
5484	carbohydrate	saccharide	1	1	1
5485	carbohydrate loading	carbohydrates	1	1	1
5486	carbohydrate	carbohydrates	1	1	1
5487	carbohydrate loading	#NAME?	1	1	1
5488	carbohydrate	#NAME?	1	1	1

5489	carbohydrate loading	polysaccharide	1	1	1
5490	carbohydrate	polysaccharide	1	1	1
5491	carbohydrate loading	deoxyribose	1	1	1
5492	carbohydrate	deoxyribose	1	1	1
5493	carbohydrate loading	glycoprotein	1	1	1
5494	carbohydrate	glycoprotein	1	1	1
5495	carbohydrate loading	jaggary	1	1	1
5496	carbohydrate	jaggary	1	1	1
5497	carbohydrate loading	jaggery	1	1	1
5498	carbohydrate	jaggery	1	1	1
5499	carbohydrate loading	jagghery	1	1	1
5500	carbohydrate	jagghery	1	1	1
5501	carbohydrate loading	monosaccharose	1	1	1
5502	carbohydrate	monosaccharose	1	1	1
5503	carbohydrate loading	oligosaccharide	1	1	1
5504	carbohydrate	oligosaccharide	1	1	1
5505	carbohydrate loading	polyose	1	1	1
5506	carbohydrate	polyose	1	1	1
5507	carbohydrate loading	agar	1	1	1
5508	carbohydrate	agar	1	1	1
5509	carbohydrate loading	amylum	1	1	1
5510	carbohydrate	amylum	1	1	1
5511	carbohydrate loading	antigen	1	1	1
5512	carbohydrate	antigen	1	1	1
5513	carbohydrate loading	arabin	1	1	1
5514	carbohydrate	arabin	1	1	1
5515	carbohydrate loading	chemosynthesis	1	1	1
5516	carbohydrate	chemosynthesis	1	1	1
5517	carbohydrate loading	citric acid	1	1	1
5518	carbohydrate	citric acid	1	1	1
5519	carbohydrate loading	comfort food	1	1	1
5520	carbohydrate	comfort food	1	1	1

5521	carbohydrate loading	cortef	1	1	1
5522	carbohydrate	cortef	1	1	1
5523	carbohydrate loading	corticosteroid	1	1	1
5524	carbohydrate	corticosteroid	1	1	1
5525	carbohydrate loading	corticosterone	1	1	1
5526	carbohydrate	corticosterone	1	1	1
5527	carbohydrate loading	cortisol	1	1	1
5528	carbohydrate	cortisol	1	1	1
5529	carbohydrate loading	dark+reaction	1	1	1
5530	carbohydrate	dark+reaction	1	1	1
5531	carbohydrate loading	dextrin	1	1	1
5532	carbohydrate	dextrin	1	1	1
5533	carbohydrate loading	fermentation	1	1	1
5534	carbohydrate	fermentation	1	1	1
5535	carbohydrate loading	galactin	1	1	1
5536	carbohydrate	galactin	1	1	1
5537	carbohydrate loading	gelose	1	1	1
5538	carbohydrate	gelose	1	1	1
5539	carbohydrate loading	glyceraldehyde	1	1	1
5540	carbohydrate	glyceraldehyde	1	1	1
5541	carbohydrate loading	glycolipid	1	1	1
5542	carbohydrate	glycolipid	1	1	1
5543	carbohydrate loading	hydrate	1	1	1
5544	carbohydrate	hydrate	1	1	1
5545	carbohydrate loading	hydrocortisone	1	1	1
5546	carbohydrate	hydrocortisone	1	1	1
5547	carbohydrate loading	hydrocortone	1	1	1
5548	carbohydrate	hydrocortone	1	1	1
5549	carbohydrate loading	lactic+acid	1	1	1
5550	carbohydrate	lactic+acid	1	1	1
5551	carbohydrate loading	lignin	1	1	1
5552	carbohydrate	lignin	1	1	1

5553	carbohydrate loading	mucoid	1	1	1
5554	carbohydrate	mucoid	1	1	1
5555	carbohydrate loading	pectose	1	1	1
5556	carbohydrate	pectose	1	1	1
5557	carbohydrate loading	phosphofructokinase	1	1	1
5558	carbohydrate	phosphofructokinase	1	1	1
5559	carbohydrate loading	refined sugar	1	1	1
5560	carbohydrate	refined sugar	1	1	1
5561	low-calorie diet	whisper	1	1	1
5562	soft	whisper	1	1	1
5563	low-calorie diet	gondola	1	1	1
5564	bland	gondola	1	1	1
5565	low-calorie diet	cocoa powder	1	1	1
5566	low-fat diet	cocoa powder	1	1	1
5567	low-calorie diet	tennis	1	1	1
5568	low-sodium diet	tennis	1	1	1
5569	low-calorie diet	elves	1	1	1
5570	low-sodium diet	elves	1	1	1
5571	low-calorie diet	kart	1	1	1
5572	low-sodium diet	kart	1	1	1
5573	low-calorie diet	burley	1	1	1
5574	low-sodium diet	burley	1	1	1
5575	low-calorie diet	cold light	1	1	1
5576	low-sodium diet	cold light	1	1	1
5577	low-calorie diet	hansom cab	1	1	1
5578	low-sodium diet	hansom cab	1	1	1
5579	low-calorie diet	tobogganing	1	1	1
5580	low-sodium diet	tobogganing	1	1	1
5581	low-calorie diet	dimmer	1	1	1
5582	low-sodium diet	dimmer	1	1	1
5583	low-calorie diet	pickup truck	1	1	1
5584	low-sodium diet	pickup truck	1	1	1
5585	low-calorie diet	proxima	1	1	1
5586	low-sodium diet	proxima	1	1	1
5587	low-calorie diet	chemiluminescence	1	1	1
5588	low-sodium diet	chemiluminescence	1	1	1
5589	low-calorie diet	leucocratic	1	1	1
5590	low-sodium diet	leucocratic	1	1	1
5591	low-calorie diet	airglow	1	1	1
5592	low-sodium diet	airglow	1	1	1
5593	low-calorie diet	toboggan	1	1	1

5594	low-sodium diet	toboggan	1	1	1
5595	low-calorie diet	phototaxy	1	1	1
5596	low-sodium diet	phototaxy	1	1	1
5597	low-calorie diet	reichsrath	1	1	1
5598	low-sodium diet	reichsrath	1	1	1
5599	low-calorie diet	virga	1	1	1
5600	low-sodium diet	virga	1	1	1
5601	low-calorie diet	oats	1	1	1
5602	low-sodium diet	oats	1	1	1
5603	low-calorie diet	sunscald	1	1	1
5604	low-sodium diet	sunscald	1	1	1
5605	low-calorie diet	coffee	1	1	1
5606	low-sodium diet	coffee	1	1	1
5607	low-calorie diet	chloroplast	1	1	1
5608	low-sodium diet	chloroplast	1	1	1
5609	spoon food	bowl	1	1	1
5610	athletic training	bowl	1	1	1
5611	low-calorie diet	thin	1	1	1
5612	liquid diet	thin	1	1	1
5613	fad diet	twist	1	1	1
5614	follow	twist	1	1	1
5615	bland	plate	1	1	1
5616	spoon food	plate	1	1	1
5617	spoon food	sup	1	1	1
5618	eat	sup	1	1	1
5619	spoon food	indulge	1	1	1
5620	eat	indulge	1	1	1
5621	spoon food	eats	1	1	1
5622	fare	eats	1	1	1
5623	spoon food	hash	1	1	1
5624	fare	hash	1	1	1
5625	fare	repeat	1	1	1
5626	athletic training	repeat	1	1	1
5627	high-protein diet	spider	1	1	1
5628	high-vitamin diet	spider	1	1	1
5629	spoon food	help	1	1	1
5630	avoid	help	1	1	1
5631	low-sodium diet	kernite	1	1	1
5632	soft diet	kernite	1	1	1
5633	soft	downy	1	1	1
5634	soft diet	downy	1	1	1
5635	soft	zephyr	1	1	1
5636	soft diet	zephyr	1	1	1

5637	soft	challis	1	1	1
5638	soft diet	challis	1	1	1
5639	lite	candle	1	1	1
5640	abstemious	candle	1	1	1
5641	reducing diet	lithium	1	1	1
5642	soft diet	lithium	1	1	1
5643	reduce	quash	1	1	1
5644	avoid	quash	1	1	1
5645	eat	breakfast	1	1	1
5646	fare	breakfast	1	1	1
5647	spare	cannibalize	1	1	1
5648	eat	cannibalize	1	1	1
5649	fare	beat	1	1	1
5650	avoid	beat	1	1	1
5651	carbo loading	overload	1	1	1
5652	carbohydrate loading	overload	1	1	1
5653	carbo loading	power loading	1	1	1
5654	carbohydrate loading	power loading	1	1	1
5655	carbo loading	stevedore	1	1	1
5656	carbohydrate loading	stevedore	1	1	1
5657	carbo loading	lighterage	1	1	1
5658	carbohydrate loading	lighterage	1	1	1
5659	carbo loading	save-all	1	1	1
5660	carbohydrate loading	save-all	1	1	1
5661	carbo loading	trainload	1	1	1
5662	carbohydrate loading	trainload	1	1	1
5663	carbo loading	autoloading	1	1	1
5664	carbohydrate loading	autoloading	1	1	1
5665	carbo loading	burthen	1	1	1
5666	carbohydrate loading	burthen	1	1	1
5667	carbo loading	cascabel	1	1	1
5668	carbohydrate loading	cascabel	1	1	1
5669	carbo loading	dead load	1	1	1
5670	carbohydrate loading	dead load	1	1	1
5671	carbo loading	fraughtage	1	1	1
5672	carbohydrate loading	fraughtage	1	1	1

5673	carbo loading	fraughting	1	1	1
5674	carbohydrate loading	fraughting	1	1	1
5675	carbo loading	hopper	1	1	1
5676	carbohydrate loading	hopper	1	1	1
5677	carbo loading	live load	1	1	1
5678	carbohydrate loading	live load	1	1	1
5679	carbo loading	millstone	1	1	1
5680	carbohydrate loading	millstone	1	1	1
5681	carbo loading	oneration	1	1	1
5682	carbohydrate loading	oneration	1	1	1
5683	carbo loading	span loading	1	1	1
5684	carbohydrate loading	span loading	1	1	1
5685	carbo loading	stevedorage	1	1	1
5686	carbohydrate loading	stevedorage	1	1	1
5687	carbo loading	superload	1	1	1
5688	carbohydrate loading	superload	1	1	1
5689	off	dispatch	1	1	1
5690	eat	dispatch	1	1	1
5691	off	crop	1	1	1
5692	eat	crop	1	1	1
5693	off	avert	1	1	1
5694	avoid	avert	1	1	1
5695	off	derail	1	1	1
5696	avoid	derail	1	1	1
5697	off	fence	1	1	1
5698	avoid	fence	1	1	1
5699	low-calorie diet	nouvelle cuisine	1	1	1
5700	low-salt diet	nouvelle cuisine	1	1	1
5701	off	spark	1	1	1
5702	light	spark	1	1	1
5703	train	jump	1	1	1
5704	off	jump	1	1	1
5705	reduce	shave	1	1	1
5706	off	shave	1	1	1
5707	balanced diet	lopsided	1	1	1
5708	off	lopsided	1	1	1
5709	reduce	burn	1	1	1
5710	eat	burn	1	1	1

5711	reduce	slenderize	1	1	1
5712	dieting	slenderize	1	1	1
5713	reduce	trim	1	1	1
5714	spare	trim	1	1	1
5715	reduce	contract	1	1	1
5716	reducing diet	contract	1	1	1
5717	reduce	decrease	1	1	1
5718	reducing diet	decrease	1	1	1
5719	reduce	reducer	1	1	1
5720	reducing diet	reducer	1	1	1
5721	hand-schuller-christian disease	pass	1	1	1
5722	spare	pass	1	1	1
5723	diabetic coma	berenice's hair	1	1	1
5724	kussmaul's coma	berenice's hair	1	1	1
5725	diabetic coma	carus	1	1	1
5726	kussmaul's coma	carus	1	1	1
5727	diabetic coma	envelope	1	1	1
5728	kussmaul's coma	envelope	1	1	1
5729	diabetic coma	hepatic coma	1	1	1
5730	kussmaul's coma	hepatic coma	1	1	1
5731	diabetic coma	semicomatose	1	1	1
5732	kussmaul's coma	semicomatose	1	1	1
5733	hand-schuller-christian disease	sign	1	1	1
5734	arbitrary	sign	1	1	1
5735	diabetic coma	comatose	1	1	1
5736	kussmaul's coma	comatose	1	1	1
5737	diabetic coma	comate	1	1	1
5738	kussmaul's coma	comate	1	1	1
5739	diabetic coma	narcoma	1	1	1
5740	kussmaul's coma	narcoma	1	1	1
5741	hand-schuller-christian disease	but	1	1	1
5742	spare	but	1	1	1

5743	hand-schuller-christian disease	christen	1	1	1
5744	schuller-christian disease	christen	1	1	1
5745	hand-schuller-christian disease	over	1	1	1
5746	schuller-christian disease	over	1	1	1
5747	hand-schuller-christian disease	crown wart	1	1	1
5748	schuller-christian disease	crown wart	1	1	1
5749	hand-schuller-christian disease	heaven	1	1	1
5750	schuller-christian disease	heaven	1	1	1
5751	hand-schuller-christian disease	nematode	1	1	1
5752	schuller-christian disease	nematode	1	1	1
5753	hand-schuller-christian disease	crown gall	1	1	1
5754	schuller-christian disease	crown gall	1	1	1
5755	hand-schuller-christian disease	molluscum	1	1	1
5756	schuller-christian disease	molluscum	1	1	1
5757	hand-schuller-christian disease	new thought	1	1	1
5758	schuller-christian disease	new thought	1	1	1
5759	hand-schuller-christian disease	agapeistic	1	1	1

5760	schuller-christian disease	agapeistic	1	1	1
5761	hand-schuller-christian disease	doctor	1	1	1
5762	schuller-christian disease	doctor	1	1	1
5763	hand-schuller-christian disease	sacrament	1	1	1
5764	schuller-christian disease	sacrament	1	1	1
5765	hand-schuller-christian disease	receive	1	1	1
5766	schuller-christian disease	receive	1	1	1
5767	hand-schuller-christian disease	kiss of peace	1	1	1
5768	schuller-christian disease	kiss of peace	1	1	1
5769	hand-schuller-christian disease	pesthouse	1	1	1
5770	schuller-christian disease	pesthouse	1	1	1
5771	hand-schuller-christian disease	erythema	1	1	1
5772	schuller-christian disease	erythema	1	1	1
5773	hand-schuller-christian disease	albigensianism	1	1	1
5774	schuller-christian disease	albigensianism	1	1	1
5775	hand-schuller-christian disease	catharism	1	1	1
5776	schuller-christian disease	catharism	1	1	1

5777	hand-schuller-christian disease	inri	1	1	1
5778	schuller-christian disease	inri	1	1	1
5779	hand-schuller-christian disease	mennonite	1	1	1
5780	schuller-christian disease	mennonite	1	1	1
5781	hand-schuller-christian disease	faith cure	1	1	1
5782	schuller-christian disease	faith cure	1	1	1
5783	hand-schuller-christian disease	kiss+of+peace	1	1	1
5784	schuller-christian disease	kiss+of+peace	1	1	1
5785	hand-schuller-christian disease	symbolism	1	1	1
5786	schuller-christian disease	symbolism	1	1	1
5787	hand-schuller-christian disease	martyrdom	1	1	1
5788	schuller-christian disease	martyrdom	1	1	1
5789	hand-schuller-christian disease	coenurus	1	1	1
5790	schuller-christian disease	coenurus	1	1	1
5791	glucose tolerance test	glycaemia	1	1	1
5792	glucose	glycaemia	1	1	1
5793	glucose tolerance test	maltase	1	1	1
5794	glucose	maltase	1	1	1
5795	glucose tolerance test	amylase	1	1	1

5796	glucose	amylase	1	1	1
5797	glucose tolerance test	corn sugar	1	1	1
5798	glucose	corn sugar	1	1	1
5799	glucose tolerance test	adrenaline	1	1	1
5800	glucose	adrenaline	1	1	1
5801	glucose tolerance test	gluconic	1	1	1
5802	glucose	gluconic	1	1	1
5803	glucose tolerance test	phosphorylase	1	1	1
5804	glucose	phosphorylase	1	1	1
5805	hypoglycaemic agent	commission	1	1	1
5806	hypoglycemic agent	commission	1	1	1
5807	glucose tolerance test	dextrose	1	1	1
5808	glucose	dextrose	1	1	1
5809	glucose tolerance test	inversion	1	1	1
5810	glucose	inversion	1	1	1
5811	glucose tolerance test	glucoside	1	1	1
5812	glucose	glucoside	1	1	1
5813	glucose tolerance test	grape sugar	1	1	1
5814	glucose	grape sugar	1	1	1
5815	glucose tolerance test	invertase	1	1	1
5816	glucose	invertase	1	1	1
5817	glucose tolerance test	dextroglucose	1	1	1
5818	glucose	dextroglucose	1	1	1
5819	glucose tolerance test	blood glucose	1	1	1
5820	glucose	blood glucose	1	1	1
5821	glucose tolerance test	blood sugar	1	1	1
5822	glucose	blood sugar	1	1	1
5823	diabetic coma	hyperalimentation	1	1	1
5824	kussmaul's coma	hyperalimentation	1	1	1
5825	diabetic coma	total parenteral nutrition	1	1	1
5826	kussmaul's coma	total parenteral nutrition	1	1	1
5827	diabetic coma	tpn	1	1	1

5828	kussmaul's coma	tpn	1	1	1
5829	diabetic coma	comae	1	1	1
5830	kussmaul's coma	comae	1	1	1
5831	diabetic coma	comatic	1	1	1
5832	kussmaul's coma	comatic	1	1	1
5833	diabetic coma	comatoseness	1	1	1
5834	kussmaul's coma	comatoseness	1	1	1
5835	diabetic coma	convulsions	1	1	1
5836	kussmaul's coma	convulsions	1	1	1
5837	diabetic coma	torpor	1	1	1
5838	kussmaul's coma	torpor	1	1	1
5839	diabetic coma	narcotic	1	1	1
5840	kussmaul's coma	narcotic	1	1	1
5841	diabetic coma	lead poisoning	1	1	1
5842	kussmaul's coma	lead poisoning	1	1	1
5843	diabetic coma	comet	1	1	1
5844	kussmaul's coma	comet	1	1	1
5845	diabetic coma	heatstroke	1	1	1
5846	kussmaul's coma	heatstroke	1	1	1
5847	diabetic coma	sleep	1	1	1
5848	kussmaul's coma	sleep	1	1	1
5849	diabetic coma	nucleus	1	1	1
5850	kussmaul's coma	nucleus	1	1	1
5851	diabetic coma	eclampsia	1	1	1
5852	kussmaul's coma	eclampsia	1	1	1
5853	diabetic coma	braxy	1	1	1
5854	kussmaul's coma	braxy	1	1	1
5855	diabetic coma	coma berenices	1	1	1
5856	kussmaul's coma	coma berenices	1	1	1
5857	diabetic coma	comose	1	1	1
5858	kussmaul's coma	comose	1	1	1
5859	diabetic coma	karen ann quinlan	1	1	1

5860	kussmaul's coma	karen ann quinlan	1	1	1
5861	diabetic coma	revive	1	1	1
5862	kussmaul's coma	revive	1	1	1
5863	diabetic coma	robin cook	1	1	1
5864	kussmaul's coma	robin cook	1	1	1
5865	diabetic coma	schmidt telescope	1	1	1
5866	kussmaul's coma	schmidt telescope	1	1	1
5867	diabetic coma	semicoma	1	1	1
5868	kussmaul's coma	semicoma	1	1	1
5869	hypoglycaemic agent	emulsifier	1	1	1
5870	hypoglycemic agent	emulsifier	1	1	1
5871	hypoglycaemic agent	engine	1	1	1
5872	hypoglycemic agent	engine	1	1	1
5873	hypoglycaemic agent	estate agent	1	1	1
5874	hypoglycemic agent	estate agent	1	1	1
5875	hypoglycaemic agent	federal official	1	1	1
5876	hypoglycemic agent	federal official	1	1	1
5877	hypoglycaemic agent	general agent	1	1	1
5878	hypoglycemic agent	general agent	1	1	1
5879	hypoglycaemic agent	house agent	1	1	1
5880	hypoglycemic agent	house agent	1	1	1
5881	hypoglycaemic agent	infection	1	1	1
5882	hypoglycemic agent	infection	1	1	1
5883	hypoglycaemic agent	intravenous anesthetic	1	1	1
5884	hypoglycemic agent	intravenous anesthetic	1	1	1
5885	hypoglycaemic agent	land agent	1	1	1
5886	hypoglycemic agent	land agent	1	1	1

5887	hypoglycaemic agent	local anaesthetic	1	1	1
5888	hypoglycemic agent	local anaesthetic	1	1	1
5889	hypoglycaemic agent	local anesthetic	1	1	1
5890	hypoglycemic agent	local anesthetic	1	1	1
5891	hypoglycaemic agent	mutagen	1	1	1
5892	hypoglycemic agent	mutagen	1	1	1
5893	hypoglycaemic agent	resolvent	1	1	1
5894	hypoglycemic agent	resolvent	1	1	1
5895	hypoglycaemic agent	spinal anaesthetic	1	1	1
5896	hypoglycemic agent	spinal anaesthetic	1	1	1
5897	hypoglycaemic agent	spinal anesthetic	1	1	1
5898	hypoglycemic agent	spinal anesthetic	1	1	1
5899	hypoglycaemic agent	steward	1	1	1
5900	hypoglycemic agent	steward	1	1	1
5901	hypoglycaemic agent	surfactant	1	1	1
5902	hypoglycemic agent	surfactant	1	1	1
5903	hypoglycaemic agent	syndic	1	1	1
5904	hypoglycemic agent	syndic	1	1	1
5905	hypoglycaemic agent	t-man	1	1	1
5906	hypoglycemic agent	t-man	1	1	1
5907	hypoglycaemic agent	teratogen	1	1	1
5908	hypoglycemic agent	teratogen	1	1	1
5909	hypoglycaemic agent	thinner	1	1	1
5910	hypoglycemic agent	thinner	1	1	1
5911	hypoglycaemic agent	topical anaesthetic	1	1	1

5912	hypoglycemic agent	topical anaesthetic	1	1	1
5913	hypoglycaemic agent	topical anesthetic	1	1	1
5914	hypoglycemic agent	topical anesthetic	1	1	1
5915	hypoglycaemic agent	active	1	1	1
5916	hypoglycemic agent	active	1	1	1
5917	hypoglycaemic agent	alcahest	1	1	1
5918	hypoglycemic agent	alcahest	1	1	1
5919	hypoglycaemic agent	alkahest	1	1	1
5920	hypoglycemic agent	alkahest	1	1	1
5921	hypoglycaemic agent	alkalizer	1	1	1
5922	hypoglycemic agent	alkalizer	1	1	1
5923	hypoglycaemic agent	antacid	1	1	1
5924	hypoglycemic agent	antacid	1	1	1
5925	hypoglycaemic agent	antidote	1	1	1
5926	hypoglycemic agent	antidote	1	1	1
5927	hypoglycaemic agent	bacteriostat	1	1	1
5928	hypoglycemic agent	bacteriostat	1	1	1
5929	hypoglycaemic agent	bailee	1	1	1
5930	hypoglycemic agent	bailee	1	1	1
5931	hypoglycaemic agent	bailiff	1	1	1
5932	hypoglycemic agent	bailiff	1	1	1
5933	hypoglycaemic agent	blanching agent	1	1	1
5934	hypoglycemic agent	blanching agent	1	1	1
5935	hypoglycaemic agent	bleaching agent	1	1	1
5936	hypoglycemic agent	bleaching agent	1	1	1

5937	hypoglycaemic agent	local	1	1	1
5938	hypoglycemic agent	local	1	1	1
5939	hypoglycaemic agent	disinfectant	1	1	1
5940	hypoglycemic agent	disinfectant	1	1	1
5941	hypoglycaemic agent	realtor	1	1	1
5942	hypoglycemic agent	realtor	1	1	1
5943	hypoglycaemic agent	bleach	1	1	1
5944	hypoglycemic agent	bleach	1	1	1
5945	hypoglycaemic agent	attorney	1	1	1
5946	hypoglycemic agent	attorney	1	1	1
5947	hypoglycaemic agent	detergent	1	1	1
5948	hypoglycemic agent	detergent	1	1	1
5949	hypoglycaemic agent	procurator	1	1	1
5950	hypoglycemic agent	procurator	1	1	1
5951	glucose tolerance test	lenient	1	1	1
5952	soft	lenient	1	1	1
5953	diabetic coma	cerebrospinal meningitis	1	1	1
5954	kussmaul's coma	cerebrospinal meningitis	1	1	1
5955	glucose tolerance test	maltose	1	1	1
5956	glucose	maltose	1	1	1
5957	diabetic coma	envelop	1	1	1
5958	kussmaul's coma	envelop	1	1	1
5959	diabetic coma	mydriasis	1	1	1
5960	kussmaul's coma	mydriasis	1	1	1
5961	glucose tolerance test	glycogenesis	1	1	1
5962	glucose	glycogenesis	1	1	1
5963	glucose tolerance test	hexose	1	1	1
5964	glucose	hexose	1	1	1

5965	glucose tolerance test	amylose	1	1	1
5966	glucose	amylose	1	1	1
5967	glucose tolerance test	galactose	1	1	1
5968	glucose	galactose	1	1	1
5969	glucose tolerance test	sucrase	1	1	1
5970	glucose	sucrase	1	1	1
5971	glucose tolerance test	lactase	1	1	1
5972	glucose	lactase	1	1	1
5973	cushing's disease	cirrhosis	1	1	1
5974	macrobiotic diet	cirrhosis	1	1	1
5975	hypoglycaemic agent	agent provocateur	1	1	1
5976	hypoglycemic agent	agent provocateur	1	1	1
5977	hypoglycaemic agent	chemical agent	1	1	1
5978	hypoglycemic agent	chemical agent	1	1	1
5979	hypoglycaemic agent	clorox	1	1	1
5980	hypoglycemic agent	clorox	1	1	1
5981	hypoglycaemic agent	coolant	1	1	1
5982	hypoglycemic agent	coolant	1	1	1
5983	hypoglycaemic agent	decoagulant	1	1	1
5984	hypoglycemic agent	decoagulant	1	1	1
5985	hypoglycaemic agent	deus ex machina	1	1	1
5986	hypoglycemic agent	deus ex machina	1	1	1
5987	hypoglycaemic agent	dissolvent	1	1	1
5988	hypoglycemic agent	dissolvent	1	1	1
5989	hypoglycaemic agent	dissolver	1	1	1
5990	hypoglycemic agent	dissolver	1	1	1
5991	hypoglycaemic agent	expectorant	1	1	1

5992	hypoglycemic agent	expectorant	1	1	1
5993	hypoglycaemic agent	fungicide	1	1	1
5994	hypoglycemic agent	fungicide	1	1	1
5995	hypoglycaemic agent	gastric antacid	1	1	1
5996	hypoglycemic agent	gastric antacid	1	1	1
5997	hypoglycaemic agent	soap	1	1	1
5998	hypoglycemic agent	soap	1	1	1
5999	hypoglycaemic agent	fed	1	1	1
6000	hypoglycemic agent	fed	1	1	1
6001	hypoglycaemic agent	g-man	1	1	1
6002	hypoglycemic agent	g-man	1	1	1
6003	hypoglycaemic agent	spiritual being	1	1	1
6004	hypoglycemic agent	spiritual being	1	1	1
6005	hypoglycaemic agent	supernatural being	1	1	1
6006	hypoglycemic agent	supernatural being	1	1	1
6007	hypoglycaemic agent	bleaching powder	1	1	1
6008	hypoglycemic agent	bleaching powder	1	1	1
6009	hypoglycaemic agent	cause	1	1	1
6010	hypoglycemic agent	cause	1	1	1
6011	hypoglycaemic agent	cautery	1	1	1
6012	hypoglycemic agent	cautery	1	1	1
6013	hypoglycaemic agent	chloride of lime	1	1	1
6014	hypoglycemic agent	chloride of lime	1	1	1
6015	hypoglycaemic agent	federal	1	1	1
6016	hypoglycemic agent	federal	1	1	1

6017	hypoglycaemic agent	germicide	1	1	1
6018	hypoglycemic agent	germicide	1	1	1
6019	hypoglycaemic agent	vasoconstrictor	1	1	1
6020	hypoglycemic agent	vasoconstrictor	1	1	1
6021	hypoglycaemic agent	whitener	1	1	1
6022	hypoglycemic agent	whitener	1	1	1
6023	hypoglycaemic agent	anticoagulant	1	1	1
6024	hypoglycemic agent	anticoagulant	1	1	1
6025	hypoglycaemic agent	antifungal	1	1	1
6026	hypoglycemic agent	antifungal	1	1	1
6027	hypoglycaemic agent	antimycotic	1	1	1
6028	hypoglycemic agent	antimycotic	1	1	1
6029	hypoglycaemic agent	business agent	1	1	1
6030	hypoglycemic agent	business agent	1	1	1
6031	hypoglycaemic agent	comprador	1	1	1
6032	hypoglycemic agent	comprador	1	1	1
6033	hypoglycaemic agent	desiccant	1	1	1
6034	hypoglycemic agent	desiccant	1	1	1
6035	hypoglycaemic agent	diluent	1	1	1
6036	hypoglycemic agent	diluent	1	1	1
6037	hypoglycaemic agent	dilutant	1	1	1
6038	hypoglycemic agent	dilutant	1	1	1
6039	iron-storage disease	puddle	1	1	1
6040	iron overload	puddle	1	1	1
6041	iron-storage disease	greensickness	1	1	1
6042	iron overload	greensickness	1	1	1

6043	iron overload	pig	1	1	1
6044	eat	pig	1	1	1
6045	iron-storage disease	bug	1	1	1
6046	cushing's disease	bug	1	1	1
6047	iron-storage disease	brand	1	1	1
6048	iron overload	brand	1	1	1
6049	iron-storage disease	hold	1	1	1
6050	well-fed	hold	1	1	1
6051	iron overload	burden	1	1	1
6052	carbo loading	burden	1	1	1
6053	carbo loading	overburden	1	1	1
6054	carbohydrate loading	overburden	1	1	1
6055	iron-storage disease	read	1	1	1
6056	well-fed	read	1	1	1
6057	iron-storage disease	terminal	1	1	1
6058	schuller-christian disease	terminal	1	1	1
6059	amino acid	penicillamine	1	1	1
6060	iron-storage disease	store	1	1	1
6061	salt-free diet	store	1	1	1
6062	iron overload	hook	1	1	1
6063	off	hook	1	1	1
6064	acetoacetic acid	acidulate	1	1	1
6065	ascorbic acid	acidulate	1	1	1
6066	ketone body	person	1	1	1
6067	ketone body	#NAME?	1	1	1
6068	ketone body	water	1	1	1
6069	ketone body	flesh	1	1	1
6070	ketone body	church	1	1	1
6071	ketone body	constituency	1	1	1
6072	ketone body	sphinx	1	1	1
6073	ketone body	ball	1	1	1
6074	iron overload	cast iron	1	1	1
6075	brittle diabetes	cast iron	1	1	1
6076	carbohydrate loading	wood sugar	1	1	1
6077	carbohydrate	wood sugar	1	1	1
6078	latent diabetes	spleen	1	1	1

6079	iron-storage disease	spleen	1	1	1
6080	diabetic diet	ary	1	1	1
6081	macrobiotic diet	ary	1	1	1
6082	diabetic diet	aschaffenburg	1	1	1
6083	macrobiotic diet	aschaffenburg	1	1	1
6084	diabetic diet	botanical medicine	1	1	1
6085	macrobiotic diet	botanical medicine	1	1	1
6086	diabetic diet	athletic training	1	1	1
6087	macrobiotic diet	athletic training	1	1	1
6088	hand-schuller-christian disease	touch	1	1	1
6089	fad diet	cult	1	1	1
6090	regimen	cult	1	1	1
6091	cushing's disease	leprosy	1	1	1
6092	hand-schuller-christian disease	leprosy	1	1	1
6093	cushing's disease	sequela	1	1	1
6094	hand-schuller-christian disease	sequela	1	1	1
6095	cushing's disease	aspergillosis	1	1	1
6096	hand-schuller-christian disease	aspergillosis	1	1	1
6097	cushing's disease	clap	1	1	1
6098	hand-schuller-christian disease	clap	1	1	1
6099	sir frederick grant banting	barrie	1	1	1
6100	john james rickard macleod	barrie	1	1	1
6101	nonketoacidosis prone	subject	1	1	1
6102	liable	subject	1	1	1
6103	nonketoacidosis prone	labile	1	1	1
6104	liable	labile	1	1	1

6105	nonketoacidosis prone	suspicious	1	1	1
6106	liable	suspicious	1	1	1
6107	nonketoacidosis prone	mutable	1	1	1
6108	liable	mutable	1	1	1
6109	nonketoacidosis prone	peccable	1	1	1
6110	liable	peccable	1	1	1
6111	nonketoacidosis prone	apt	1	1	1
6112	liable	apt	1	1	1
6113	nonketoacidosis prone	likely	1	1	1
6114	liable	likely	1	1	1
6115	cushing's disease	syphilis	1	1	1
6116	hand-schuller-christian disease	syphilis	1	1	1
6117	hypoglycaemic agent	emmenagogue	1	1	1
6118	hypoglycemic agent	emmenagogue	1	1	1
6119	antidiabetic drug	prescription	1	1	1
6120	botanical medicine	prescription	1	1	1
6121	antidiabetic drug	painkiller	1	1	1
6122	botanical medicine	painkiller	1	1	1
6123	antidiabetic drug	placebo	1	1	1
6124	botanical medicine	placebo	1	1	1
6125	antidiabetic drug	astringent	1	1	1
6126	botanical medicine	astringent	1	1	1
6127	antidiabetic drug	pharmaceutical	1	1	1
6128	botanical medicine	pharmaceutical	1	1	1
6129	antidiabetic drug	physic	1	1	1
6130	botanical medicine	physic	1	1	1
6131	antidiabetic drug	antiviral	1	1	1

6132	botanical medicine	antiviral	1	1	1
6133	antidiabetic drug	pain pill	1	1	1
6134	botanical medicine	pain pill	1	1	1
6135	sir frederick grant banting	fleming	1	1	1
6136	john james rickard macleod	fleming	1	1	1
6137	sir frederick grant banting	murray	1	1	1
6138	john james rickard macleod	murray	1	1	1
6139	sir frederick grant banting	hall	1	1	1
6140	john macleod	hall	1	1	1
6141	sir frederick grant banting	morgan	1	1	1
6142	john macleod	morgan	1	1	1
6143	sir frederick grant banting	wallace	1	1	1
6144	charles herbert best	wallace	1	1	1
6145	sir frederick grant banting	ross	1	1	1
6146	john james rickard macleod	ross	1	1	1
6147	sir frederick grant banting	thomson	1	1	1
6148	john james rickard macleod	thomson	1	1	1
6149	sir frederick grant banting	grey	1	1	1
6150	charles herbert best	grey	1	1	1
6151	soft diet	fluorite	1	1	1
6152	soft diet	fluorspar	1	1	1
6153	pancreas	pancreatic	1	1	1
6154	glucagon	pancreatic	1	1	1
6155	pancreas	islets	1	1	1
6156	isles of langerhans	islets	1	1	1
6157	brittle diabetes	bismuth	1	1	1
6158	sir frederick grant banting	weld	1	1	1
6159	frederick sanger	weld	1	1	1
6160	sir frederick grant banting	dewar	1	1	1

6161	frederick sanger	dewar	1	1	1
6162	sir frederick grant banting	ashton	1	1	1
6163	frederick sanger	ashton	1	1	1
6164	sir frederick grant banting	hopkins	1	1	1
6165	frederick sanger	hopkins	1	1	1
6166	sir frederick grant banting	pollock	1	1	1
6167	frederick sanger	pollock	1	1	1
6168	sir frederick grant banting	whitaker	1	1	1
6169	frederick sanger	whitaker	1	1	1
6170	sir frederick grant banting	ondine	1	1	1
6171	frederick sanger	ondine	1	1	1
6172	sir frederick grant banting	herschel	1	1	1
6173	frederick sanger	herschel	1	1	1
6174	sir frederick grant banting	north	1	1	1
6175	frederick sanger	north	1	1	1
6176	sir frederick grant banting	wilkins	1	1	1
6177	frederick sanger	wilkins	1	1	1
6178	sir frederick grant banting	barbarossa	1	1	1
6179	frederick sanger	barbarossa	1	1	1
6180	sir frederick grant banting	frederick the great	1	1	1
6181	frederick sanger	frederick the great	1	1	1
6182	sir frederick grant banting	frederiksberg	1	1	1
6183	frederick sanger	frederiksberg	1	1	1
6184	sir frederick grant banting	handel	1	1	1
6185	frederick sanger	handel	1	1	1
6186	ketone body	ketose	1	1	1
6187	ketone body	ketosteroid	1	1	1
6188	ketone body	keto	1	1	1
6189	ketone body	anthraquinone	1	1	1
6190	ketone body	testosterone	1	1	1
6191	ketone body	ketohexose	1	1	1
6192	ketone body	flavin	1	1	1
6193	ketone body	butyrone	1	1	1
6194	ketone body	laurone	1	1	1
6195	ketone body	margarone	1	1	1

6196	ketone body	myristone	1	1	1
6197	ketone body	oenanthone	1	1	1
6198	ketone body	palmitone	1	1	1
6199	ketone body	propione	1	1	1
6200	ketone body	suberone	1	1	1
6201	ketone body	valerone	1	1	1
6202	ketone body	xanthone	1	1	1
6203	ketone body	flavanone	1	1	1
6204	ketone body	ketonic	1	1	1
6205	ketone body	benzoin	1	1	1
6206	ketone body	stearone	1	1	1
6207	ketone body	benzophenone	1	1	1
6208	ketone body	dimethyl ketone	1	1	1
6209	ketone body	hexone	1	1	1
6210	ketone body	thienone	1	1	1
6211	ketone body	acetophenone	1	1	1
6212	ketone body	methyl isobutyl ketone	1	1	1
6213	ketone body	androsterone	1	1	1
6214	ketone body	camphor	1	1	1
6215	ketone body	pregnenolone	1	1	1
6216	ketone body	benzoquinone	1	1	1
6217	ketone body	butanone	1	1	1
6218	ketone body	ketoxime	1	1	1
6219	ketone body	methyl ethyl ketone	1	1	1
6220	ketone body	cinnamone	1	1	1
6221	ketone body	oleone	1	1	1
6222	ketone body	quinone	1	1	1
6223	ketone body	ketone group	1	1	1
6224	ketone body	chassis	1	1	1
6225	ketone body	state	1	1	1
6226	ketone body	transit	1	1	1
6227	ketone body	passage	1	1	1
6228	ketone body	somatic	1	1	1
6229	ketone body	anatomy	1	1	1
6230	ketone body	frame	1	1	1
6231	ketone body	orb	1	1	1
6232	ketone body	college	1	1	1
6233	ketone body	corpus	1	1	1
6234	ketone body	figure	1	1	1
6235	ketone body	government	1	1	1
6236	ketone body	member	1	1	1
6237	ketone body	shape	1	1	1
6238	ketone body	system	1	1	1
6239	ketone body	adult body	1	1	1
6240	ketone body	carcass	1	1	1

6241	ketone body	incorporate	1	1	1
6242	ketone body	physique	1	1	1
6243	ketone body	skin	1	1	1
6244	ketone body	soma	1	1	1
6245	ketone body	temperature	1	1	1
6246	ketone body	carbonyl	1	1	1
6247	ketone body	hemiacetal	1	1	1
6248	ketone body	clay	1	1	1
6249	ketone body	corselet	1	1	1
6250	type i diabetes	impression	1	1	1
6251	produced	impression	1	1	1
6252	gestational diabetes	child	1	1	1
6253	juvenile	child	1	1	1
6254	gestational diabetes	kid	1	1	1
6255	juvenile	kid	1	1	1
6256	gestational diabetes	fry	1	1	1
6257	juvenile	fry	1	1	1
6258	gestational diabetes	nestling	1	1	1
6259	juvenile	nestling	1	1	1
6260	gestational diabetes	youngster	1	1	1
6261	juvenile	youngster	1	1	1
6262	gestational diabetes	younker	1	1	1
6263	juvenile	younker	1	1	1
6264	gestational diabetes	nipper	1	1	1
6265	juvenile	nipper	1	1	1
6266	gestational diabetes	shaver	1	1	1
6267	juvenile	shaver	1	1	1
6268	gestational diabetes	small fry	1	1	1
6269	juvenile	small fry	1	1	1
6270	gestational diabetes	juvenility	1	1	1
6271	juvenile	juvenility	1	1	1
6272	gestational diabetes	tike	1	1	1
6273	juvenile	tike	1	1	1
6274	gestational diabetes	tyke	1	1	1
6275	juvenile	tyke	1	1	1

6276	gestational diabetes	minor	1	1	1
6277	juvenile	minor	1	1	1
6278	gestational diabetes	tiddler	1	1	1
6279	juvenile	tiddler	1	1	1
6280	gestational diabetes	young person	1	1	1
6281	juvenile	young person	1	1	1
6282	gestational diabetes	delinquent	1	1	1
6283	juvenile	delinquent	1	1	1
6284	gestational diabetes	juvenile delinquency	1	1	1
6285	juvenile	juvenile delinquency	1	1	1
6286	gestational diabetes	ingenue	1	1	1
6287	juvenile	ingenue	1	1	1
6288	gestational diabetes	juvenile delinquent	1	1	1
6289	juvenile	juvenile delinquent	1	1	1
6290	gestational diabetes	delinquency	1	1	1
6291	juvenile	delinquency	1	1	1
6292	gestational diabetes	hebesphalmology	1	1	1
6293	juvenile	hebesphalmology	1	1	1
6294	gestational diabetes	juvenile court	1	1	1
6295	juvenile	juvenile court	1	1	1
6296	gestational diabetes	apple box	1	1	1
6297	juvenile	apple box	1	1	1
6298	type i diabetes	coordinate	1	1	1
6299	type ii diabetes	coordinate	1	1	1
6300	type i diabetes	generation	1	1	1
6301	type ii diabetes	generation	1	1	1
6302	type i diabetes	stick	1	1	1
6303	iron overload	stick	1	1	1
6304	type i diabetes	perfect	1	1	1
6305	mature-onset diabetes	perfect	1	1	1
6306	type i diabetes	throw	1	1	1
6307	off	throw	1	1	1
6308	low-fat diet	chip	1	1	1
6309	off	chip	1	1	1
6310	type i diabetes	logotype	1	1	1

6311	type ii diabetes	logotype	1	1	1
6312	type i diabetes	division	1	1	1
6313	type ii diabetes	division	1	1	1
6314	type i diabetes	diiodide	1	1	1
6315	type ii diabetes	diiodide	1	1	1
6316	type i diabetes	pithecanthropus	1	1	1
6317	type ii diabetes	pithecanthropus	1	1	1
6318	type i diabetes	face	1	1	1
6319	type ii diabetes	face	1	1	1
6320	type i diabetes	lead	1	1	1
6321	type ii diabetes	lead	1	1	1
6322	type i diabetes	space	1	1	1
6323	type ii diabetes	space	1	1	1
6324	adult-onset diabetes	woman	1	1	1
6325	adult-onset diabetes mellitus	woman	1	1	1
6326	adult-onset diabetes	imago	1	1	1
6327	adult-onset diabetes mellitus	imago	1	1	1
6328	adult-onset diabetes	bull	1	1	1
6329	adult-onset diabetes mellitus	bull	1	1	1
6330	adult-onset diabetes	stag	1	1	1
6331	adult-onset diabetes mellitus	stag	1	1	1
6332	adult-onset diabetes	tenor	1	1	1
6333	adult-onset diabetes mellitus	tenor	1	1	1
6334	adult-onset diabetes	bombing	1	1	1
6335	mature-onset diabetes	bombing	1	1	1
6336	adult-onset diabetes	horse	1	1	1
6337	adult-onset diabetes mellitus	horse	1	1	1
6338	adult-onset diabetes	eve	1	1	1

6339	adult-onset diabetes mellitus	eve	1	1	1
6340	adult-onset diabetes	madam	1	1	1
6341	adult-onset diabetes mellitus	madam	1	1	1
6342	adult-onset diabetes	warmonger	1	1	1
6343	adult-onset diabetes mellitus	warmonger	1	1	1
6344	adult-onset diabetes	banzai charge	1	1	1
6345	mature-onset diabetes	banzai charge	1	1	1
6346	adult-onset diabetes	teen	1	1	1
6347	adult-onset diabetes mellitus	teen	1	1	1
6348	adult-onset diabetes	boy	1	1	1
6349	adult-onset diabetes mellitus	boy	1	1	1
6350	amino acid	homology	1	1	1
6351	type i diabetes	edition	1	1	1
6352	spare	edition	1	1	1
6353	maturity-onset diabetes	autumn	1	1	1
6354	maturity-onset diabetes mellitus	autumn	1	1	1
6355	maturity-onset diabetes	haggard	1	1	1
6356	maturity-onset diabetes mellitus	haggard	1	1	1
6357	maturity-onset diabetes	indehiscent	1	1	1
6358	maturity-onset diabetes mellitus	indehiscent	1	1	1
6359	maturity-onset diabetes	maturate	1	1	1
6360	maturity-onset diabetes mellitus	maturate	1	1	1

6361	maturity-onset diabetes	prime of life	1	1	1
6362	maturity-onset diabetes mellitus	prime of life	1	1	1
6363	maturity-onset diabetes	ripeness	1	1	1
6364	maturity-onset diabetes mellitus	ripeness	1	1	1
6365	maturity-onset diabetes	rising	1	1	1
6366	maturity-onset diabetes mellitus	rising	1	1	1
6367	growth-onset diabetes	bud	1	1	1
6368	mature-onset diabetes	bud	1	1	1
6369	growth-onset diabetes	seed	1	1	1
6370	mature-onset diabetes	seed	1	1	1
6371	sir frederick grant banting	call	1	1	1
6372	off	call	1	1	1
6373	pancreas	pancreatic juice	1	1	1
6374	islets of langerhans	pancreatic juice	1	1	1
6375	pancreas	pancreatin	1	1	1
6376	islets of langerhans	pancreatin	1	1	1
6377	chemical diabetes	cf	1	1	1
6378	pancreas	cf	1	1	1
6379	banting	banteng	1	1	1
6380	banting	bos banteng	1	1	1
6381	banting	tsine	1	1	1
6382	banting	bantingism	1	1	1
6383	sugar diabetes	diastase	1	1	1
6384	pancreas	diastase	1	1	1
6385	sugar diabetes	amylopsin	1	1	1
6386	pancreas	amylopsin	1	1	1
6387	maturity-onset diabetes	grown-up	1	1	1
6388	maturity-onset diabetes mellitus	grown-up	1	1	1

6389	maturity-onset diabetes	ripe	1	1	1
6390	maturity-onset diabetes mellitus	ripe	1	1	1
6391	maturity-onset diabetes	womanhood	1	1	1
6392	maturity-onset diabetes mellitus	womanhood	1	1	1
6393	maturity-onset diabetes	coccus	1	1	1
6394	maturity-onset diabetes mellitus	coccus	1	1	1
6395	maturity-onset diabetes	emerging	1	1	1
6396	maturity-onset diabetes mellitus	emerging	1	1	1
6397	non-insulin-dependent diabetes	independent	1	1	1
6398	non-insulin-dependent diabetes mellitus	independent	1	1	1
6399	non-insulin-dependent diabetes	contingent	1	1	1
6400	non-insulin-dependent diabetes mellitus	contingent	1	1	1
6401	non-insulin-dependent diabetes	anaclitic	1	1	1
6402	non-insulin-dependent diabetes mellitus	anaclitic	1	1	1
6403	non-insulin-dependent diabetes	adjective	1	1	1
6404	non-insulin-dependent diabetes mellitus	adjective	1	1	1
6405	non-insulin-dependent diabetes	varistor	1	1	1

6406	non-insulin-dependent diabetes mellitus	varistor	1	1	1
6407	non-insulin-dependent diabetes	white man's burden	1	1	1
6408	non-insulin-dependent diabetes mellitus	white man's burden	1	1	1
6409	non-insulin-dependent diabetes	free	1	1	1
6410	non-insulin-dependent diabetes mellitus	free	1	1	1
6411	non-insulin-dependent diabetes	conditional	1	1	1
6412	non-insulin-dependent diabetes mellitus	conditional	1	1	1
6413	non-insulin-dependent diabetes	un-	1	1	1
6414	non-insulin-dependent diabetes mellitus	un-	1	1	1
6415	non-insulin-dependent diabetes	conventional	1	1	1
6416	non-insulin-dependent diabetes mellitus	conventional	1	1	1
6417	non-insulin-dependent diabetes	contingency	1	1	1
6418	non-insulin-dependent diabetes mellitus	contingency	1	1	1
6419	non-insulin-dependent diabetes	relative	1	1	1
6420	non-insulin-dependent	relative	1	1	1

	diabetes mellitus				
6421	maturity-onset diabetes	growth	1	1	1
6422	maturity-onset diabetes mellitus	growth	1	1	1
6423	maturity-onset diabetes	mature	1	1	1
6424	maturity-onset diabetes mellitus	mature	1	1	1
6425	non-insulin-dependent diabetes	implicit	1	1	1
6426	latent diabetes	implicit	1	1	1
6427	non-insulin-dependent diabetes	secondary	1	1	1
6428	non-insulin-dependent diabetes mellitus	secondary	1	1	1
6429	ketone body	satellite	1	1	1
6430	non-insulin-dependent diabetes	precarious	1	1	1
6431	liable	precarious	1	1	1